



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**
FACULTAD DE EDUCACIÓN TÉCNICA
PARA EL DESARROLLO
CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

TEMA:

**Implementación de dos arquitecturas de redes neuronales
para la clasificación de cáncer de mamas**

AUTOR:

Romero Bersosa, Mayra Alejandra

Componente práctico del examen complejo previo a la
obtención del título de **INGENIERA EN
TELECOMUNICACIONES**

REVISOR:

Palacios Meléndez, Edwin Fernando

Guayaquil, Ecuador

17 de Marzo del 2017



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

FACULTAD DE EDUCACIÓN TÉCNICA PARA EL DESARROLLO
CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

CERTIFICACIÓN

Certificamos que el presente **componente práctico del examen complejo**, fue realizado en su totalidad por **Romero Bersosa, Mayra Alejandra** como requerimiento para la obtención del título de **INGENIERA EN TELECOMUNICACIONES**.

REVISOR

Palacios Meléndez, Edwin Fernando

DIRECTOR DE CARRERA

Heras Sánchez, Miguel Armando

Guayaquil, a los 17 del mes de Marzo del año 2017



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

FACULTAD DE EDUCACIÓN TÉCNICA PARA EL DESARROLLO
CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

DECLARACIÓN DE RESPONSABILIDAD

Yo, **Romero Bersoja, Mayra Alejandra**

DECLARÓ QUE:

El **componente práctico del examen complejo, “Implementación de dos arquitecturas de redes neuronales para la clasificación de cáncer de mamas”** previo a la obtención del Título de **Ingeniera en Telecomunicaciones**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Guayaquil, a los 17 del mes de Marzo del año 2017

EL AUTOR

ROMERO BERSOZA, MAYRA ALEJANDRA



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

FACULTAD DE EDUCACIÓN TÉCNICA PARA EL DESARROLLO
CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

AUTORIZACIÓN

Yo, **Romero Bersoja, Mayra Alejandra**

Autorizó a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del **componente práctico del examen complejo, “Implementación de dos arquitecturas de redes neuronales para la clasificación de cáncer de mamas”**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

Guayaquil, a los 17 del mes de Marzo del año 2017

EL AUTOR

ROMERO BERSOZA, MAYRA ALEJANDRA

REPORTE DE URKUND

The screenshot shows the URKUND interface. On the left, document details are listed: **Documento**: Romero_Alejandra_EC_Final.docx (D26578861); **Presentado**: 2017-03-20 07:37 (-05:00); **Presentado por**: fernandopm23@hotmail.com; **Recibido**: edwin.palacios.ucsg@analysis.orkund.com; **Mensaje**: Revisión Componente Práctico de Alejandra Romero. A yellow highlight indicates that 2% of the document's text is derived from 3 sources. On the right, a 'Lista de fuentes' (List of sources) table is visible, listing various documents and URLs. At the bottom right, a yellow warning box states '1 Advertencia' (1 Warning).

Lista de fuentes	Bloques		
+	Categoría	Enlace/nombre de archivo	
+		Vera y Swett.doc	<input type="checkbox"/>
+		Diana-Guaman.docx	<input type="checkbox"/>
+	>	Protocolo de cáncer de mama.docx	<input type="checkbox"/>
+		TRABAJO FINAL 1 marzo 2015.docx	<input type="checkbox"/>
+		http://www.canceronline.cl/index.php?opti...	<input checked="" type="checkbox"/>

UNIVERSIDAD CATÓLICA DE SANTIAGO DE GUAYAQUIL
FACULTAD DE EDUCACIÓN TÉCNICA PARA EL DESARROLLO
CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

TEMA: Implementación de dos arquitecturas de redes neuronales para la clasificación de cáncer de mamas

AUTOR: Romero Bersosa, Mayra Alejandra

Componente práctico del examen complejo previo a la obtención del grado de INGENIERA EN TELECOMUNICACIONES

REVISOR: Palacios Meléndez, Edwin Fernando

Guayaquil, Ecuador

13 de Marzo del 2017

DEDICATORIA

A mis padres, Maritza Bersoza y Jorge Romero, por el amor incondicional que me han dado y la paciencia por cumplir con la meta de ser Ingeniera en Telecomunicaciones.

EL AUTOR

ROMERO BERSOZA, MAYRA ALEJANDRA

AGRADECIMIENTO

A mis padres, por la paciencia y apoyo condicional que me han brindado.

A Dios, por estar conmigo en esta etapa finalizada.

A la Universidad Católica de Santiago de Guayaquil, por haber contribuido en mi formación académica y personal.

EL AUTOR

ROMERO BERSOZA, MAYRA ALEJANDRA



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL**

FACULTAD DE EDUCACIÓN TÉCNICA PARA EL DESARROLLO
CARRERA DE INGENIERÍA EN TELECOMUNICACIONES

TRIBUNAL DE SUSTENTACIÓN

f. _____

CORDOVA RIVADENEIRA, LUIS SILVIO
REVISOR

f. _____

ROMERO PAZ, MANUEL DE JESUS
DECANO

f. _____

PALACIOS MELÉNDEZ, EDWIN FERNANDO
COORDINADOR DE TITULACIÓN

Índice General

Índice de Figuras	XI
Resumen	XIII
CAPÍTULO 1: Descripción general del componente práctico	2
1.1. Introducción.....	2
1.2. Justificación del problema de investigación.....	3
1.3. Definición del problema de investigación.....	3
1.4. Definición del problema de investigación.....	4
1.5. Objetivos de investigación.....	4
1.5.1. Objetivo general.....	4
1.5.2. Objetivos específicos:	4
1.6. Hipótesis del componente práctico.....	4
1.7. Metodología de investigación.....	5
CAPÍTULO 2: Fundamentación Teórica.....	6
2.1. Fundamentos del diagnóstico asistido por computador	6
2.2. Funcionamiento general del diagnóstico asistido por computador	8
2.3. Procesos que componen los CAD.....	8
2.4. Redes neuronales artificiales	10
2.5. Tipos de aprendizajes	15
2.6. Aprendizaje supervisado: algoritmo Backpropagation	16
2.7. Overfitting.....	18
CAPÍTULO 3: Implementación de varias arquitecturas para clasificación de cáncer de mamas.....	23
3.1. Base de datos utilizada	23
3.2. Experimentos	24
3.3. Detalles de la implementación.....	25
3.4. Resultados obtenidos.....	31

Conclusiones	36
Referencias Bibliográficas	37

Índice de Figuras

Capítulo 2

Figura 2. 1: Modelo no lineal de una neurona artificial.	11
Figura 2. 2: Función de activación lineal: "hardlim".	13
Figura 2. 3: Función de activación lineal: "hardlims".	13
Figura 2. 4: Función de activación lineal: "pureline".	14
Figura 2. 5: Función de activación no lineal: "sigmoide"	14
Figura 2. 6: Función de activación no lineal: "tangente hiperbólica"	15
Figura 2. 7: Flujo de procesamiento del algoritmo backpropagation	17
Figura 2. 8: Sistema de predicción de dos clases en que la línea verde representa un modelo sobreajustado y la línea negra un modelo regularizado.	19
Figura 2. 9: Curva del error de entrenamiento y validación. Eje vertical: error, eje horizontal: tiempo.	20

Capítulo 3

Figura 3. 1: Arquitecturas de perceptron multicapas implementadas (a) 3-10-1 (b) 5-10-1	24
Figura 3. 2: (a) Parámetros de configuración (criterio 1) (b) Curva de convergencia (Arq. 3-10-1)	26
Figura 3. 3: (a) Parámetros de configuración (criterio 1) (b) Curva de convergencia (Arq. 5-10-1)	27
Figura 3. 4: (a) Parámetros de configuración (criterio 2) (b) Curva de convergencia (Arq. 3-10-1)	28
Figura 3. 5: (a) Parámetros de configuración (criterio 2) (b) Curva de convergencia (Arq. 5-10-1)	29
Figura 3. 6: (a) Parámetros de configuración (criterio 3) (b) Curva de convergencia (Arq. 3-10-1)	30
Figura 3. 7: (a) Parámetros de configuración (criterio 3) (b) Curva de convergencia (Arq. 5-10-1)	30

Figura 3. 8: Matriz de confusión. Criterio: MSE (a) Arq. 3-10-1 (b) Arq. 5-10-1	31
Figura 3. 9: Matriz de confusión. Criterio: Parada anticipada (a) Arq. 3-10-1 (b) Arq. 5-10-1	32
Figura 3. 10: Matriz de confusión. Criterio: Regularización bayesiana (a) Arq. 3-10-1 (b) Arq. 5-10-1	33

Resumen

El cáncer de mama es el cáncer más frecuente en la mujer y representa la primera causa de muerte en países desarrollados y en vías de desarrollo. Su incidencia está aumentando en el mundo debido a la mayor esperanza de vida, el aumento de la urbanización y la adopción de modos de vida occidentales. La identificación del cáncer de mama depende de la interpretación del médico a partir de las informaciones obtenidas de los pacientes a través de exámenes. Las tecnologías para el diagnóstico asistido por computadora dan soporte a los médicos en la interpretación de los contenidos multimedia obtenidos en pruebas a las que se ha sometido el paciente. Las redes neuronales artificiales son fundamentales dentro de estos sistemas. Ellas son un modelo artificial y simplificado del cerebro humano, capaces de adquirir conocimiento a través de la experiencia. En este trabajo son implementadas dos arquitecturas de redes neuronales para la clasificación de biopsias mamarias en malignas y benignas. Las arquitecturas fueron entrenadas con tres criterios de parada cada una, los cuales son: error cuadrático medio, parada anticipada y regularización bayesiana. Los mejores valores de exactitud promedio se obtuvieron para las redes entrenadas con el criterio de parada anticipada.

Palabras claves: DIAGNÓSTICO ASISTIDO POR COMPUTADORA, PRE-PROCESAMIENTO, REDES NEURONALES ARTIFICIALES, CÁNCER DE MAMA, OVERFITTING, BACKPROPAGATION.

CAPÍTULO 1: Descripción general del componente práctico

1.1. Introducción

A nivel mundial, el cáncer de mama es la primera causa de muerte por cáncer en la mujer tanto en los países desarrollados como en los países en desarrollo. Según la Organización Mundial de la Salud, la incidencia de cáncer de mama está aumentando en el mundo debido a la mayor esperanza de vida, el aumento de la urbanización y la adopción de modos de vida occidentales (“Ministerio de Salud - Ecuador,” 2017). Esta enfermedad consiste en el crecimiento desordenado de células del tejido de la mama, formando nódulos que pueden ser malignos o benignos. Las investigaciones clínicas han demostrado que el diagnóstico precoz aumenta las tasas de supervivencia en pacientes con cáncer de mama (Shapiro S, Venet W, Strax P, Venet L, 1982).

Se estima que anualmente fallecen 458 mil mujeres por esta causa. La tasa de incidencia (número de casos nuevos) es de 39 casos por 100 000 mujeres. Se estima que anualmente fallecen 23 000 mujeres por cáncer de mama en el continente americano (“Portal - Cancer Online,” 2017). En Ecuador el cáncer de mama es uno de los más frecuentes entre las mujeres, 30 de cada 100 000 personas son diagnosticadas con esta enfermedad cada año y casi la mitad de ellas en etapas avanzadas de la enfermedad, lo cual disminuye las posibilidades de curación y supervivencia (“Ministerio de Salud - Ecuador,” 2017).

La identificación del cáncer de mama depende de la interpretación del médico a partir de las informaciones obtenidas de los pacientes a través de exámenes, los cuales incluyen el examen clínico, la mamografía y el análisis del tejido de la mama. Generalmente, cuando existen lesiones sospechosas, el médico busca confirmación del diagnóstico de forma citológica, por medio de punción aspirativa por aguja fina (PAAF), o histológica, cuando el material

es obtenido por punción, utilizándose aguja gruesa (PAG) o biopsia quirúrgica convencional.

Las tecnologías para el diagnóstico asistido por computadora (CAD, del inglés *Computer Aided Diagnosis*), dan soporte a los médicos en la interpretación de los contenidos multimedia obtenidos en pruebas a las que se ha sometido el paciente, como por ejemplo las imágenes médicas. También son una ayuda para la toma de decisiones. La idea de estos sistemas no es la de sustituir la labor del médico ni la de dar un diagnóstico completo sino servir como un complemento o “segunda opinión” y como ayuda para conseguir un diagnóstico óptimo (Doi, 2007).

1.2. Justificación del problema de investigación.

Los sistemas CAD son una tecnología interdisciplinaria, que combina la inteligencia artificial, el procesamiento digital de imágenes, el reconocimiento de patrones y la radiología. El procesamiento de las imágenes o del video, basado en complejos sistemas de reconocimiento de patrones permite que el especialista, generalmente un radiólogo, pueda interpretar la información con mucha menos dificultad. La subjetividad propia del diagnóstico visual puede ser minimizada con el análisis digital de la imagen basado en técnica de aprendizaje de máquina (más conocido por su término en inglés, *machine learning*).

1.3. Definición del problema de investigación.

El análisis de redes neuronales en la Carrera de Ingeniería en Telecomunicaciones no ha sido tratado ni abordado en trabajos ni en proyectos de investigación. El procesamiento de señales digitales permite modelar diferentes señales inclusive las biomédicas, por ejemplo, la clasificación de biopsias mamarias en malignas o benignas.

1.4. Definición del problema de investigación.

Las redes neuronales son un modelo artificial y simplificado del cerebro humano, que es capaz de adquirir conocimiento a través de la experiencia. Su unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona. Dichas redes permiten clasificar datos con un entrenamiento previo. Este enfoque puede auxiliar a los profesionales de la salud en el procesamiento de las informaciones generadas por los exámenes de laboratorio citopatológicos de cáncer de mama, de forma a llegar a una conclusión con respecto a la observación analizada, sea maligno o benigno, en el caso de nódulos mamarios.

1.5. Objetivos de investigación.

1.5.1. Objetivo general

Implementación de dos arquitecturas de redes neuronales artificiales entrenadas con tres criterios de parada y comparación de su desempeño en la clasificación de biopsias mamarias.

1.5.2. Objetivos específicos:

- Estudiar los fundamentos teóricos de los sistemas de diagnóstico asistido por computador.
- Estudiar los fundamentos teóricos de las redes neuronales artificiales.
- Implementar en Matlab dos arquitecturas de redes neuronales artificiales entrenadas con tres criterios de parada cada una.
- Comparar el desempeño de cada experimento en la clasificación de biopsias mamarias.

1.6. Hipótesis del componente práctico.

Si se implementa una arquitectura de red neuronal artificial debidamente entrenada, se podrá clasificar las muestras de entrada como benignas o malignas.

1.7. Metodología de investigación.

El componente práctico del examen complejo, utiliza un enfoque totalmente cuantitativo, cuyo método consiste en la simulación de redes neuronales a través de MatLab. El tipo de investigación es descriptivo, ya que consiste en fundamentar teóricamente los sistemas de diagnóstico asistido por computador y de las redes neuronales artificiales. También es explicativo, ya que a través del diseño realizado de dos arquitecturas de redes neuronales.

CAPÍTULO 2: Fundamentación Teórica

En este capítulo se abordará la fundamentación teórica que soporta el presente trabajo. Serán expuestos los fundamentos teóricos de los sistemas de diagnóstico asistido por computador. Serán introducidas las redes neuronales artificiales, su unidad básica: la neurona, sus principios de funcionamiento y los métodos de aprendizaje de la red. También se explicará el algoritmo *backpropagation*, que es el núcleo del entrenamiento y las estrategias contra el sobreajuste de la red.

2.1. Fundamentos del diagnóstico asistido por computador

El diagnóstico asistido por computador puede ser definido como un diagnóstico hecho por un radiólogo que utiliza el resultado de análisis cuantitativos automatizados de imágenes radiográficas como una “segunda opinión” para la toma de decisiones diagnósticas. Es importante resaltar que la computadora es usada solamente como una herramienta para la obtención de información adicional, pues el diagnóstico final es dado por el especialista.

La finalidad de los CAD es mejorar la exactitud del diagnóstico, así como la consistencia de la interpretación de la imagen radiológica, mediante el uso de la respuesta del computador como referencia. La respuesta del computador puede ser útil, una vez que el diagnóstico del especialista está basado en una evaluación subjetiva que está sujeta a variaciones intra e inter personales, así como pérdida de información debido a la calidad de la imagen, sobreposición de estructuras, fatiga visual o distracción. Además de eso, fue demostrado que una lectura doble (o sea, por dos especialistas) puede aumentar la sensibilidad del diagnóstico. La propuesta del CAD es funcionar como un segundo especialista (Azevedo-Marquez, 2001).

Básicamente, existen dos tipos de aplicaciones de sistemas CAD. Uno es el apoyo a la detección de lesiones, a partir de la localización de patrones

anormales a través del barrido de la imagen por la computadora (por ejemplo, agrupamientos de microcalcificaciones en imágenes mamográficas o nódulos pulmonares en imágenes del tórax). El otro es el auxilio al diagnóstico, a través de la cuantificación de características de la imagen y su clasificación como correspondiendo a patrones normales o anormales (por ejemplo, la asociación de la cantidad y forma de las microcalcificaciones presentes en un agrupamiento con la malignidad o no del tumor, o la asociación de la textura de los pulmones con lesiones intersticiales en imágenes del tórax).

En general, en los sistemas CAD se utilizan técnicas provenientes de dos áreas del conocimiento: la visión computacional, que comprende el procesamiento de imagen para su realce, la segmentación y la extracción de características, e la inteligencia artificial, que incluye métodos para la selección de características y reconocimiento de patrones (Azevedo-Marquez, 2001).

La base conceptual de los CAD es genérica y amplia, lo cual permite que ellos sean aplicados a todas las modalidades de obtención de imagen, incluyendo la radiografía convencional, la tomografía computarizada, la resonancia magnética, la ultrasonografía y la medicina nuclear. También se pueden desarrollar esquemas de CADs para todos los tipos de exámenes de todas las partes del cuerpo, como cráneo, tórax, abdomen, huesos y sistema vascular, entre otros.

Sin embargo, los principales objetos de investigación en el desarrollo de sistemas CAD han sido las áreas de mamografía, para la detección precoz del cáncer de mama; tórax, para la detección de nódulos pulmonares, lesiones intersticiales y neumotórax; y angiografía, para el análisis cuantitativo de estenosis y de flujo sanguíneo (Azevedo-Marquez, 2001).

2.2. Funcionamiento general del diagnóstico asistido por computador

Los sistemas CAD emplean algoritmos para reconocer patrones en datos de pacientes que pueden indicar posibles anomalías. De la misma forma que a un clínico se le enseña a identificar anomalías estudiando casos, a los métodos subyacentes de los CAD se les entrena para reconocer patrones en una base de datos finita que contiene casos anómalos y normales. A esta base de datos se le conoce como “conjunto de entrenamiento”.

Después que el sistema CAD ha sido entrenado, ya está listo para utilizarse en datos de nuevos pacientes para detectar patrones de anomalías parecidas a las del entrenamiento. Las clasificaciones de patrones en sistemas CAD deben ser lo suficientemente confiables y eficientes para ayudar al especialista en la identificación y valoración de las patologías. En general, la confiabilidad se estima mediante una base de datos diferente a la del entrenamiento, conocida como “conjunto de prueba”.

2.3. Procesos que componen los CAD.

Pre-procesamiento

En este primer paso, se corrigen las imperfecciones de las imágenes adquiridas, se aplican filtros para disminuir el ruido. También se armoniza la imagen en caso de sufrir diferencias en los niveles de intensidad de los píxeles en distintos puntos, o sea, se mejora el contraste de la imagen por medio de la ecualización del histograma.

Segmentación

Consiste en agrupar píxeles que pertenezcan a un mismo objeto o región subdividiendo la imagen en sus partes u objetos constituyentes. El objetivo es delinear estructuras importantes que posteriormente se analizaran individualmente. Esta etapa es fundamental para el éxito de los pasos posteriores.

Extracción de características

La extracción de características es, básicamente, un proceso de extracción de la información útil contenida en la señal a tratar (en este caso nos estamos refiriendo a imágenes). Este proceso permite la descripción de la imagen digitalizada por medio de la cuantificación de determinadas propiedades, conocidas como características. Cada región segmentada se analiza para extraer informaciones importantes, como por ejemplo: forma, tamaño, perímetro, compacidad, textura, relación con estructuras cercanas, etc. De esta forma, las regiones quedan caracterizadas con informaciones cuantitativas útiles para el diagnóstico.

Dependiendo de su naturaleza, las características extraídas pueden o no ser compatibles con la visión humana. Es posible la definición de un gran número de atributos basados en fórmulas matemáticas que pueden o no ser fácilmente comprendidos por un observador humano. Sin embargo, generalmente es más útil definir, particularmente en las etapas iniciales de desarrollo de sistemas CAD, atributos de imagen que ya hayan sido anteriormente reconocidos y descritos subjetivamente por los radiólogos, una vez que estos están basados en la observación de un número grande de casos y posibilitan diagnósticos con alta exactitud y confiabilidad (Azevedo-Marquez, 2001).

Selección de características

Este paso es opcional. Ocurre que en las estructuras de clasificación, introducir un gran número de características no siempre implica un mejor desempeño. Las técnicas de selección de características escogen aquellas con menor valor de correlación, o sea, busca quedarse con las más diferentes entre sí.

Clasificación

Luego de la extracción y selección de características, estos valores cuantificados son utilizados en el proceso de clasificación de las estructuras

presentes en la imagen. La inteligencia artificial supone el uso de la computadora para el procesamiento de datos para distinguir patrones normales y anormales, a partir de los atributos extraídos de las imágenes (Azevedo-Marquez, 2001).

Los valores numéricos resultantes de la etapa anterior sirven como entrada a los algoritmos de clasificación. Algunos ejemplos de algoritmos de clasificación son: regla del vecino más próximo, conocido como k-NN (*k-nearest neighbors*), clasificador de distancia mínima, clasificador en cascada, clasificador bayesiano de Naive, máquinas de vectores de soporte, redes neuronales artificiales (RNA), entre otros. Estas últimas son el tipo de clasificador utilizado en el presente trabajo.

2.4. Redes neuronales artificiales

Las redes neuronales artificiales son sistemas de procesamiento paralelo cuyo funcionamiento está inspirado en la estructura física del cerebro humano. Entre las múltiples ventajas del uso de las redes neuronales en la computación, se pueden citar las siguientes (Ryman-Tubb, 1993):

- Capacidad de aprendizaje (a través de ejemplos): Las RNA aprenden con la experiencia, buscando perfeccionar su desempeño y adaptarse a los escenarios nuevos y dinámicos (generalización);
- Robustez: habilidad para tratar con ruidos. Son tolerantes a las fallas y pueden presentar degradación gradual, o sea, a pesar de existir alguna falla en el sistema, continúan ofreciendo respuestas adecuadas por un período significativo de tiempo;
- Velocidad de procesamiento: pueden operar a velocidades considerables en comparación a los métodos computacionales comunes debido a que las unidades procesan de forma paralela.

Las RNA son una herramienta computacional muy utilizada para la solución de problemas complejos en aplicaciones del mundo real. En muchos casos su desempeño ha sido considerado superior a los métodos estadísticos (Falas, 1995). La capacidad de aprendizaje adaptativo es una de las características más atractivas de las redes neuronales. Esto es, aprenden a llevar a cabo ciertas tareas mediante un entrenamiento con ejemplos ilustrativos. Como las redes neuronales pueden aprender a diferenciar patrones mediante ejemplos y entrenamientos, no es necesario elaborar modelos a priori ni necesidad de especificar funciones de distribución de probabilidad.

Las redes neuronales son sistemas dinámicos auto-adaptativos. Son adaptables debido a la capacidad de autoajuste de los elementos procesales (neuronas) que componen el sistema. Son dinámicos, pues son capaces de estar constantemente en cambio para adaptarse a las nuevas condiciones. Una neurona artificial puede entenderse como una unidad de procesamiento sencilla desde el punto de vista matemático, que recibe una o más entradas, transformándolas en salidas. Cada entrada tiene un peso asociado, que determina su intensidad (Ryman-Tubb, 1993). En la figura 2.1 se presenta el modelo de una neurona artificial.

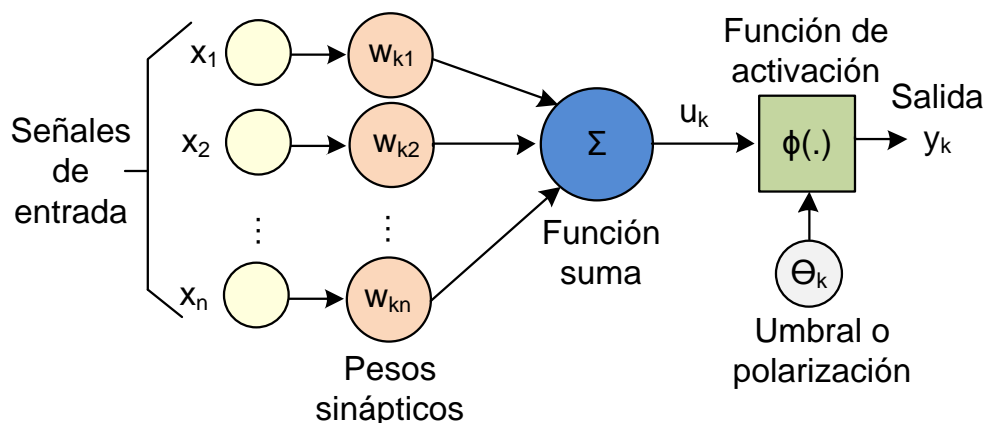


Figura 2. 1: Modelo no lineal de una neurona artificial.

Fuente: (Haykin, 1998)

Elaborado por: Autor

Las neuronas que reciben estímulos del exterior son llamadas neuronas de entrada, las que se comunican con el medio externo son llamadas neuronas de salida y las demás son conocidas como neuronas intermedias u ocultas (hidden). La capa intermedia le otorga a la red la capacidad de resolver problemas donde las clases a clasificar no son linealmente separables.

La neurona trata a muchos valores de entrada como si fueran uno solo; esto recibe el nombre de entrada global. Las entradas simples $x_1, x_2 \dots x_n$ de la Figura 2. 1: se combinan a través de la función de entrada de la siguiente forma para formar la entrada global u_k :

$$u_k = (x_1 \cdot w_{k1} + x_2 \cdot w_{k2} + \dots + x_n \cdot w_{kn}) \quad (1)$$

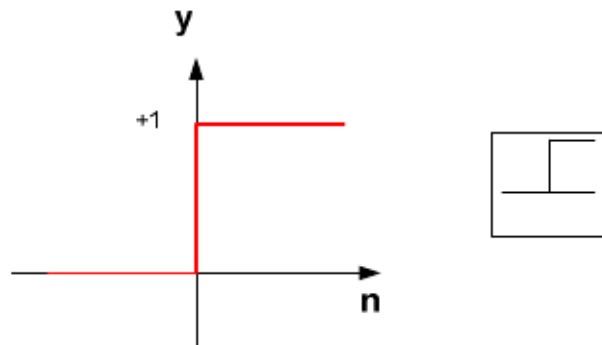
donde $x_1, x_2 \dots x_n$ son los elementos del vector de entrada, $w_{k1}, w_{k2} \dots w_{kn}$ son los pesos con son ponderadas cada una de las entradas individuales.

Los valores de entrada se multiplican por los pesos anteriormente ingresados a la neurona. Por consiguiente, los pesos que generalmente no están restringidos cambian la medida de influencia que tienen los valores de entrada. Es decir, que permiten que un gran valor de entrada tenga solamente una pequeña influencia, si estos son lo suficientemente pequeños.

Una neurona biológica puede estar activa (excitada) o inactiva (no excitada); es decir, que tiene un “estado de activación”. Las neuronas artificiales también tienen diferentes estados de activación; algunas de ellas solamente dos, al igual que las biológicas, pero otras pueden tomar cualquier valor dentro de un conjunto determinado.

La función de activación, representada por el bloque $\phi(\cdot)$ de la figura 2.1, transforma la entrada global sumado al valor de polarización $\phi(u_k + \Theta_k)$

en un valor (estado) de activación, cuyo rango normalmente va de (0 a 1) o de (-1 a 1), en el caso de funciones lineales como las mostradas en la figuras 2.2 y 2.3. Existe otra función de activación lineal, mostrada en la figura 2.4, llamada *pureline*. Ella pone a la salida el mismo valor de la entrada.

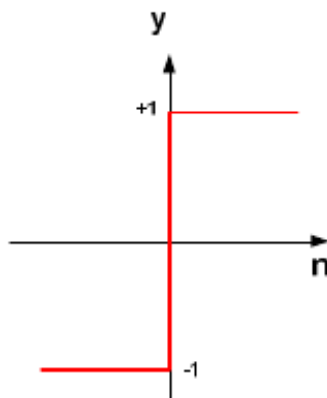


$$y = \text{hardlim}(n)$$

$$\begin{cases} y = +1 & \text{se } n \geq 0 \\ y = 0, & \text{se } n < 0 \end{cases}$$

Figura 2. 2: Función de activación lineal: “*hardlim*”.

Elaborado por: Autor



$$y = \text{hardlims}(n)$$

$$\begin{cases} y = +1 & \text{se } n \geq 0 \\ y = -1, & \text{se } n < 0 \end{cases}$$

Figura 2. 3: Función de activación lineal: “*hardlims*”.

Elaborado por: Autor

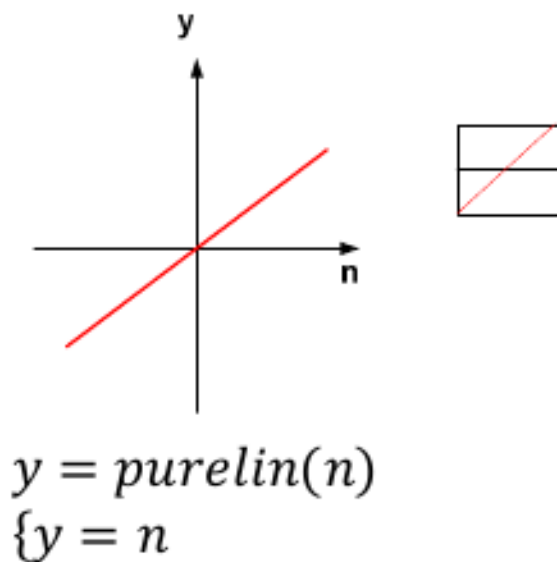


Figura 2. 4: Función de activación lineal: "pureline".
Elaborado por: Autor

También existen funciones de activación no lineales, muy usadas en la práctica, ellas son la sigmoide y la tangente hiperbólica, las mismas que son mostradas en las figuras 2.5 y 2.6 con los nombres del Matlab.

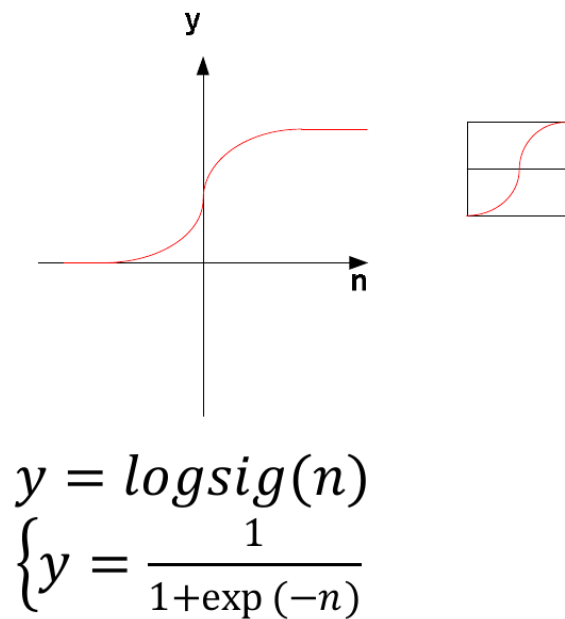
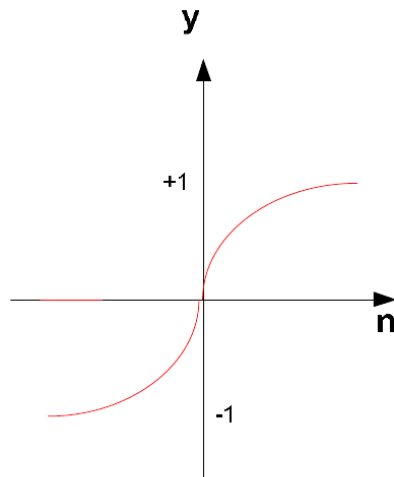


Figura 2. 5: Función de activación no lineal: "sigmoide"
Elaborado por: Autor



$$y = \text{tansig}(n)$$

$$\left\{ y = \frac{\exp(n) - \exp(-n)}{\exp(n) + \exp(-n)} \right.$$

Figura 2. 6: Función de activación no lineal: “tangente hiperbólica”
Elaborado por: Autor

Los valores de salida de la función sigmoide están comprendidos dentro de un rango que va de 0 a 1 y los de la función tangente hiperbólica entre -1 y 1.

2.5. Tipos de aprendizajes

En el proceso de aprendizaje, los enlaces ponderados de las neuronas se ajustan de manera que se obtengan ciertos resultados específicos. Una red neuronal no necesita un algoritmo para resolver un problema, ya que ella puede generar su propia distribución de pesos en los enlaces mediante el aprendizaje. La función del diseñador es únicamente la obtención de la arquitectura apropiada. No es problema del diseñador el cómo la red aprenderá a discriminar. Sin embargo, sí es necesario que desarrolle un buen algoritmo de aprendizaje que le proporcione a la red la capacidad de discriminar, mediante un entrenamiento con patrones.

El aprendizaje, o entrenamiento, es un proceso iterativo en el cual los pesos, o sea las ponderaciones de las conexiones de la red neuronal, red neuronal son ajustados. La manera particular de ajustar los pesos define el tipo específico de aprendizaje. Los métodos de aprendizaje pueden dividirse en:

- Aprendizaje supervisado: posee un “supervisor” externo que proporciona las entradas y salidas deseadas de la red, además de ajustar los pesos de las conexiones a fin de minimizar el error. La medida de error se calcula comparando la salida obtenida con la salida deseada. Es el tipo de aprendizaje usado en el presente trabajo.
- Aprendizaje no supervisado: Como el nombre indica, no posee “supervisor”, solamente la red recibe los patrones de entrada y aprende internamente a identificar y agrupar en clases.

2.6. Aprendizaje supervisado: algoritmo *Backpropagation*

El algoritmo *backpropagation* es el algoritmo de aprendizaje supervisado más conocido para el entrenamiento de redes neuronales multicapas. Este nombre resulta de la forma en que el error es propagado hacia atrás a través de la red neuronal, en otras palabras, el error se propaga hacia atrás desde la capa de salida. Esto permite que los pesos sobre las conexiones de las neuronas ubicadas en las capas ocultas cambien durante el entrenamiento. Para esto utiliza pares de datos (patrón de entrada, salida deseada).

El entrenamiento ocurre en dos fases, que recorren sentidos opuestos: *forward* y *backward*, como se muestra en la figura 2.7. En la fase *forward* la red estima la salida a partir de un patrón de entrada, de acuerdo a como estén ajustados los pesos en ese momento. Durante el entrenamiento cada dato que se presente en la entrada sigue el sentido *forward* de la entrada para la salida.

La fase *backward* usa la salida estimada por la red y la salida deseada para ajustar los pesos de las conexiones. La diferencia entre la salida

estimada y la deseada constituye el error. El objetivo del proceso de entrenamiento es disminuir cada vez más el valor de este error, por eso el valor de los pesos debe ser ajustado en cada nueva iteración. Los errores siguen el sentido *backward* de la salida para la entrada (Haykin, 1998).

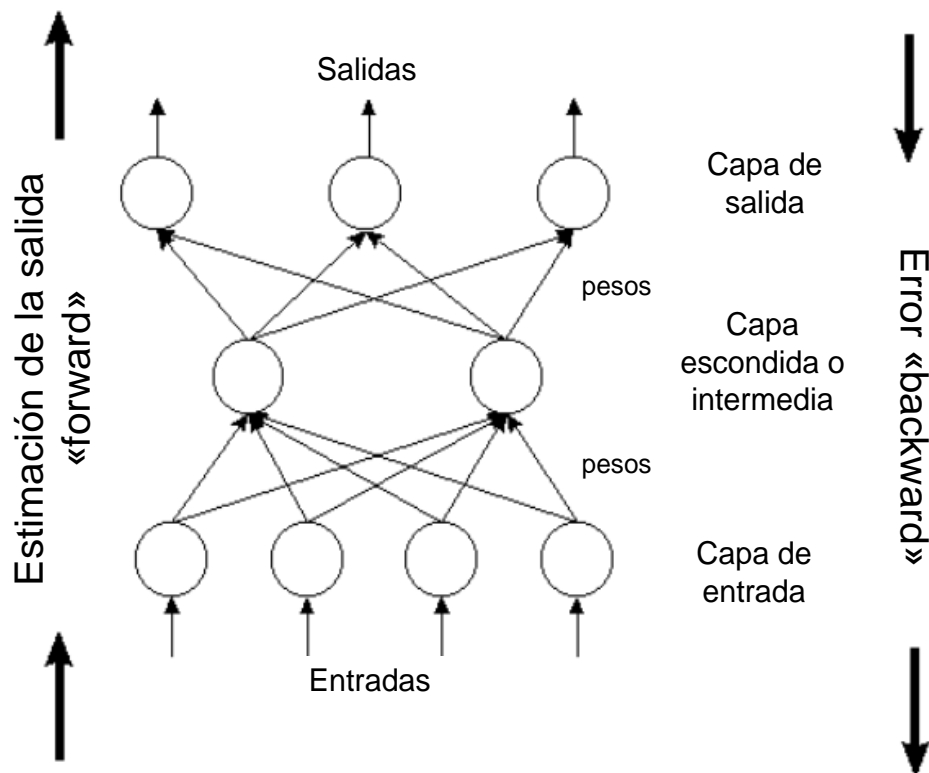


Figura 2. 7: Flujo de procesamiento del algoritmo *backpropagation*
 Fuente: (Caudill & Butler, 1992)
 Elaborado por: Autor

El algoritmo *backpropagation* está basado en la regla delta de Widrow-Hoff, y por eso puede ser llamado de regla delta generalizada. A través de este algoritmo, es posible ajustar los pesos de las capas intermedias a través del método del gradiente, con el objetivo de minimizar el error entre la salida de la red y la salida deseada (Haykin, 1998).

La función de costo a ser minimizada con la derivación de la regla delta generalizada se define, según la ecuación 2, como la suma de los errores cuadráticos:

$$E = \frac{1}{2} \sum_p \sum_{i=1}^k (d_i^p - y_i^p)^2 \quad (2)$$

donde: E es la medida del error total de la red, p es el número de patrones a clasificar, k es el número de unidades de salida, d_i es la i -ésima salida deseada, y_i es la i -ésima salida generada por la red.

Durante el entrenamiento, esta función de costo se minimiza y los pesos de las neuronas quedan ajustados para que cuando se presente un patrón en la entrada, la respuesta estimada por la red sea lo más próxima posible al valor real que se espera.

2.7. Overfitting

Una red demasiado entrenada puede que no funcione bien con datos que nunca haya “visto”. A este fenómeno se le conoce con el nombre de *overfitting* o generalización pobre y ocurre cuando una red neuronal se especializa en los datos con los que fue entrenada perdiendo así la capacidad de generalización, o sea de trabajar con nuevos datos. En ese caso, aunque el error en el conjunto de entrenamiento tenga como resultado un valor muy pequeño, cuando nuevos datos son presentados a la red, el error tiene a aumentar considerablemente (D. J. MacKay, 1992).

En la figura 2.8 se ilustra un ejemplo gráfico que ayuda a entender este concepto. Los puntos rojos pertenecen a una clase y los azules a otra, se desea que la red neuronal clasifique a que clase pertenece un determinado patrón de entrada. Si para clasificar, la red sigue la línea verde, se va a adaptar tanto a esos patrones que ante nuevos datos va a presentar grandes errores. En el caso de la clasificación siguiendo la línea negra, la red va a poder generalizar más.

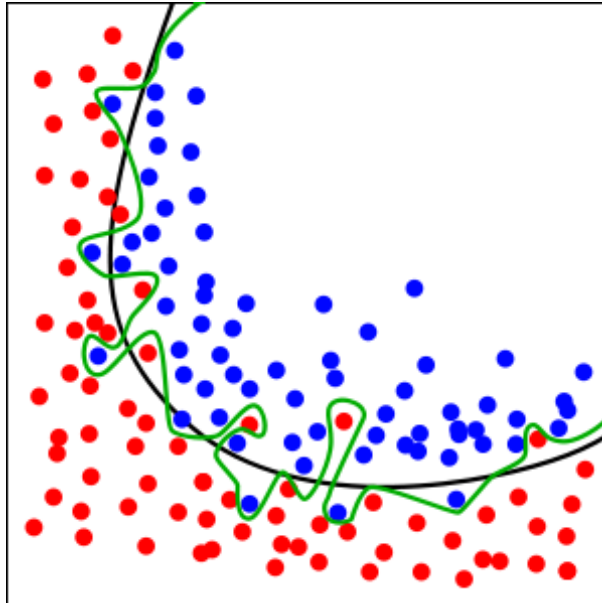


Figura 2. 8: Sistema de predicción de dos clases en que la línea verde representa un modelo sobreajustado y la línea negra un modelo regularizado.

Fuente: (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014)

Para evitar el *overfitting* se debe entrenar la red con suficientes patrones, o sea en grandes conjuntos de entrenamiento. También existen varios métodos que han sido propuestos para darle solución a este problema, entre los cuales veremos dos: la parada anticipada, en que el entrenamiento se termina en cuanto surge *overtraining*, lo cual pasa cuando la precisión de la previsión de la red entrenada, aplicada en el conjunto de prueba, empeora. La segunda técnica es la regularización bayesiana, que minimiza el problema del *overfitting* llevando en consideración la arquitectura de la red (Doan & Liong, 2004).

Parada Anticipada (*Early Stopping*)

En este método se requiere que el conjunto de datos sea dividido en tres subconjuntos: entrenamiento, validación y prueba. El primero se usa para calcular el gradiente y actualizar los pesos de la red. El segundo sirve para monitorizar el error durante el entrenamiento de la red. El error de validación normalmente disminuye durante la fase inicial de entrenamiento, así como el error del conjunto de entrenamiento. Sin embargo, cuando la red

comienza a “super ajustar” (en inglés *overfit*) los datos, el error en el conjunto de validación típicamente comienza a subir. Cuando el error de validación aumenta por un determinado número de iteraciones, definido por el usuario, el entrenamiento se interrumpe, y se retorna a los pesos con el cual se obtuvo el mínimo error de validación. En la figura 2.9 están representados los errores de entrenamiento y validación y es posible observar el momento en que el error de validación aumenta.

El tercer subconjunto es el de prueba y se utiliza para comprobar la exactitud del modelo entrenado. Este no es usado durante el entrenamiento de la red, pero si para medir el desempeño y comparar con otros modelos. También es útil medir el error del conjunto de prueba durante el proceso de entrenamiento. Si el error en el conjunto de prueba alcanza un mínimo en un número de iteración significativamente diferente de lo que lo hace el conjunto de validación, eso puede indicar una inadecuada división del conjunto de datos (Doan & Liong, 2004).

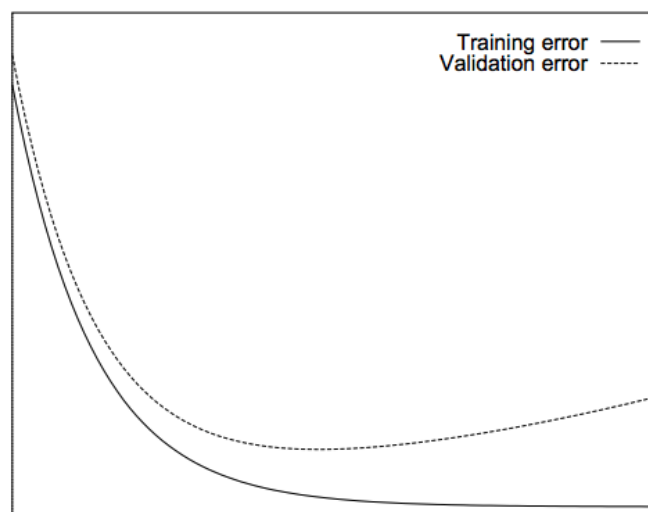


Figura 2. 9: Curva del error de entrenamiento y validación. Eje vertical: error, eje horizontal: tiempo.

Elaborado por: (Prechelt, 1998)

En resumen, la principal característica de este método es que cuando el error cuadrático medio del conjunto de validación crece un determinado número de iteraciones seguidas, se finaliza el entrenamiento, o sea, en fases

posteriores del proceso de entrenamiento, la exactitud del modelo puede comenzar a empeorar para el conjunto de validación. Esta es la fase en que el modelo debe dejar de ser entrenado para superar el problema del *overfitting*.

Regularización bayesiana

La regularización constituye un criterio de parada de entrenamiento alternativo a la parada anticipada analizada anteriormente. El método de regularización bayesiana implica modificar la función de costo, que normalmente es el error cuadrático medio (MSE, del inglés *Mean Square Error*).

El error cuadrático medio sobre el conjunto de datos de prueba (MSE_d) se expresa mediante una suma de cuadrados de los errores individuales e_i de cada muestra de un conjunto de prueba de N datos de la siguiente forma:

$$MSE_d = \frac{1}{N} \sum_{i=1}^N (e_i)^2 \quad (3)$$

Esta modificación mencionada anteriormente tiene el propósito de mejorar la capacidad de generalización del modelo de red neuronal, o sea, lograr una mayor adaptabilidad de la red a los diversos casos que se le pueden presentar. Para esto, la función de costo de la ecuación 3 se amplía con la adición de un término MSE_w que incluye el efecto de la suma de cuadrados de los pesos de la red, es decir:

$$MSE_w = \frac{1}{N} \sum_{i=1}^N (e_w)^2 \quad (4)$$

Y la función de costo modificada queda definida de la siguiente forma:

$$MSE = \beta \cdot MSE_d + \alpha \cdot MSE_w \quad (5)$$

Donde β y α son parámetros que deben ser ajustados según la metodología Bayesiana de MacKay (MacKay, 1991). Esta metodología

asume que los pesos y polarizaciones son variables aleatorias que siguen distribuciones específicas (normalmente gaussianas). Los parámetros de la regularización están relacionados con las varianzas desconocidas asociadas a estas distribuciones y se puede estimar mediante técnicas estadísticas. Una característica de este algoritmo es que proporciona una medida de cuantos parámetros de la red (pesos y polarizaciones) están siendo efectivamente usados por esta. Este número de parámetros efectivos debería llegar a ser aproximadamente constante, sin importar cuan grande se haga el número de parámetros “reales” de la red. Para esto se asume que la red ha sido entrenada durante un gran número de iteraciones y que el entrenamiento ha convergido.

La regularización bayesiana funciona mejor cuando tanto el rango de entrada como el de salida se encuentran entre $[-1, 1]$. Si no se da el caso para un determinado problema, es conveniente normalizar los datos.

CAPÍTULO 3: Implementación de varias arquitecturas para clasificación de cáncer de mamas.

En este capítulo se describen las características de la base de datos utilizada. También se explica la metodología seguida para la implementación de dos arquitecturas de redes neuronales y los criterios de parada usados. Finalmente se exponen los resultados obtenidos con cada una de las configuraciones en la clasificación de biopsias mamarias en benignas o malignas.

3.1. Base de datos utilizada

- *Wisconsin Diagnostic Breast Cancer (WDBC)*

Esta base de datos creada por Wolberg, Street y Mangasarian (Wolberg, Street, & Mangasarian, 1994) proporciona las características de los núcleos de células de tejido mamario obtenidas mediante punciones aspirativas con aguja fina de 569 pacientes, de las cuales 212 son de tejido mamario con cáncer y 357 son de mamas normales. Todos los tipos de cáncer y algunas de las masas benignas fueron histológicamente confirmados.

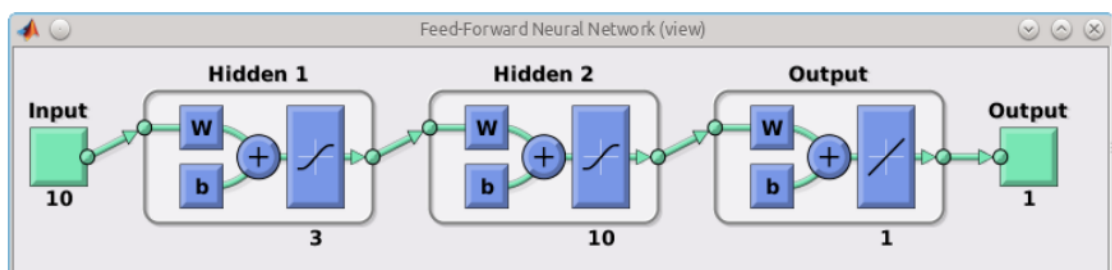
La adquisición de las imágenes de las células del tejido mamario obtenidas por la PAAF fue realizada con una resolución de 512 x 480 píxeles y 8 *bits* de profundidad. El contorno final del núcleo de la célula fue segmentado a partir de la aplicación del método de contornos activos, más conocido como *snakes*.

De cada núcleo segmentado se extrajeron 10 características diferentes: radio, área, perímetro, textura, suavidad, compacidad, concavidad, puntos cóncavos, simetría, dimensión fractal. De cada una fueron calculados el valor promedio, el peor valor (promedio de los tres mayores valores) y el error estándar. Las características extraídas fueron modeladas de tal forma que los valores más elevados están asociados a malignidad.

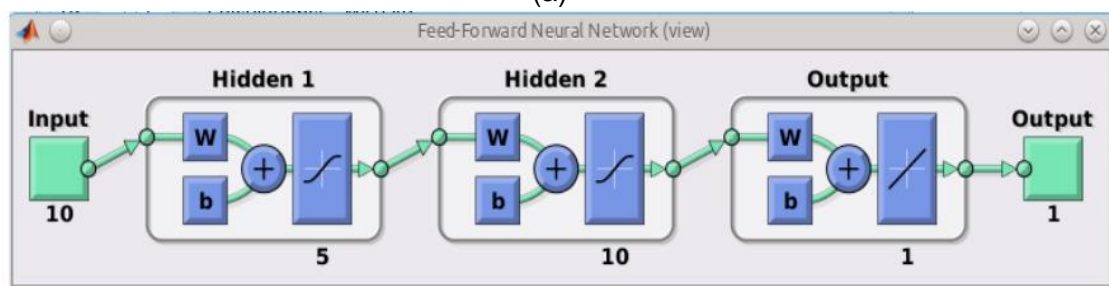
Cada línea de la base de datos representa uno de los 569 pacientes, siendo el primer valor el índice de la muestra, el segundo valor es el diagnóstico: maligno (M) o benigno (B). Después los siguientes 10 valores pertenecen a los valores promedios de las características, los próximos 10 a los errores estándar y los últimos 10 al peor valor.

3.2. Experimentos

En este trabajo son implementadas dos arquitecturas de redes neuronales (perceptron multicapas), según se muestra en la figura 3.1. En ambas la entrada es un vector columna con 10 elementos. Esos 10 elementos son las 10 características de cada biopsia. La primera arquitectura está formada por dos capas escondidas, una de 3 neuronas y otra de 10, y la salida. La segunda está formada por dos capas escondidas, una de 5 neuronas y otra de 10, y la salida. Cada una de ellas fue entrenada de acuerdo a tres criterios de parada: (1) Error cuadrático medio, (2) Parada anticipada, y (3) Regularización bayesiana.



(a)



(b)

Figura 3. 1: Arquitecturas de perceptron multicapas implementadas (a) 3-10-1 (b) 5-10-1

Elaborador por: Autor

3.3. Detalles de la implementación

Variables de entrada:

Las variables de entrada de la red neuronal son los valores promedios de cada una de las características extraídas. Como son 10 características por cada núcleo celular, tendremos 10 valores de entrada por cada muestra.

Salidas esperadas:

Las muestras fueron asociadas a un vector de salidas esperadas, definiendo 0 para el diagnóstico de “maligno” y 1 para “benigno”.

Para la creación de las redes neuronales multicapas se utilizó la función *feedforwardnet* del Matlab. Esta función permite la creación de una serie de capas, donde la primera de ellas tiene una conexión desde los datos de entrada. Cada capa posterior tiene sus neuronas conectadas a todas las de la capa anterior. La capa final produce la salida de la red, o sea, el resultado de la clasificación.

Estas redes multicapas realizan transformaciones no lineales sobre los datos y se pueden utilizar para cualquier tipo de mapeo entrada - salida. Una red con al menos una capa oculta y suficientes neuronas en dicha capa, puede adaptarse a cualquier problema donde se necesite mapear una función de entrada y producir una determinada salida.

- Criterio de parada #1: Error cuadrático medio $< 0,01$ o 10 000 iteraciones.

Con este criterio de parada, la red será entrenada hasta alcanzar un valor de MSE menor que 0,01 o hasta alcanzar las 10 000 iteraciones, lo que ocurra primero. En la figura 3.2 y figura 3.3 se muestran los parámetros de configuración y las curvas de convergencia para las arquitecturas 3-10-1 y 5-10-1 respectivamente. El método de aceleración seleccionado fue el Levenberg – Marquardt que minimiza el MSE, el cual es el algoritmo que minimiza la función de costo.

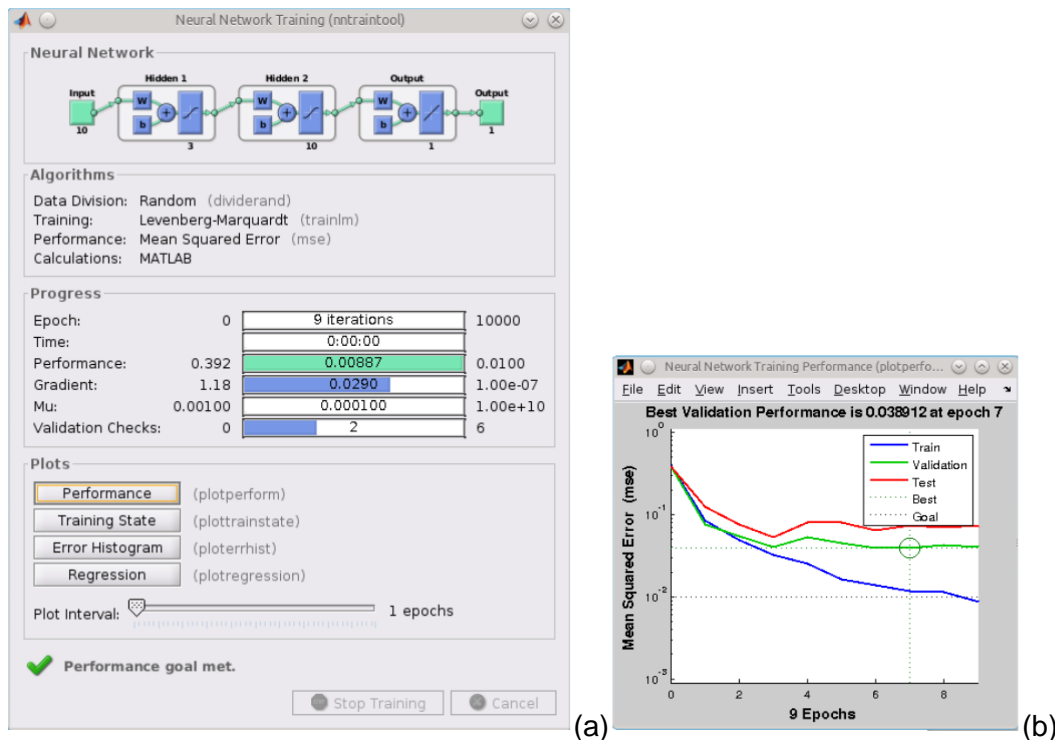


Figura 3. 2: (a) Parámetros de configuración (criterio 1) (b) Curva de convergencia (Arq. 3-10-1)

Elaborado por: Autor

En este caso, los datos fueron divididos aleatoriamente mitad y mitad para entrenamiento y prueba. Para esto se utilizó la función *dividerand*, que recibe un vector con el total de patrones y lo divide en tres subconjuntos utilizando índices aleatorios. No obstante, para este criterio de parada solo fueron creados dos subconjuntos a partir de las 569 muestras de biopsias. Los valores entrados a la función fueron 0,5 y 0,5 para entrenamiento y prueba respectivamente, o sea una división del 50 % para cada uno.

En el Matlab se definieron los siguientes parámetros:

Número de épocas: 10 000 \rightarrow *net.trainParam.epochs* = 10000;

MSE: 10^{-2} \rightarrow *net.trainParam.goal* = $1e^{-2}$

La configuración anterior significa que la red va a estar entrenando hasta el error, o sea, la función de pérdida, disminuya hasta 10^{-2} .

En ocasiones la red no logra ajustar los pesos de las neuronas de la forma más adecuada para lograr pequeñas pérdidas y sucede que la red se mantiene iterando infinitamente sin lograr la convergencia. Para prevenir esto, si se alcanzan las 10000 iteraciones y no se ha obtenido un valor de error de 10^{-2} , la red se detendrá.

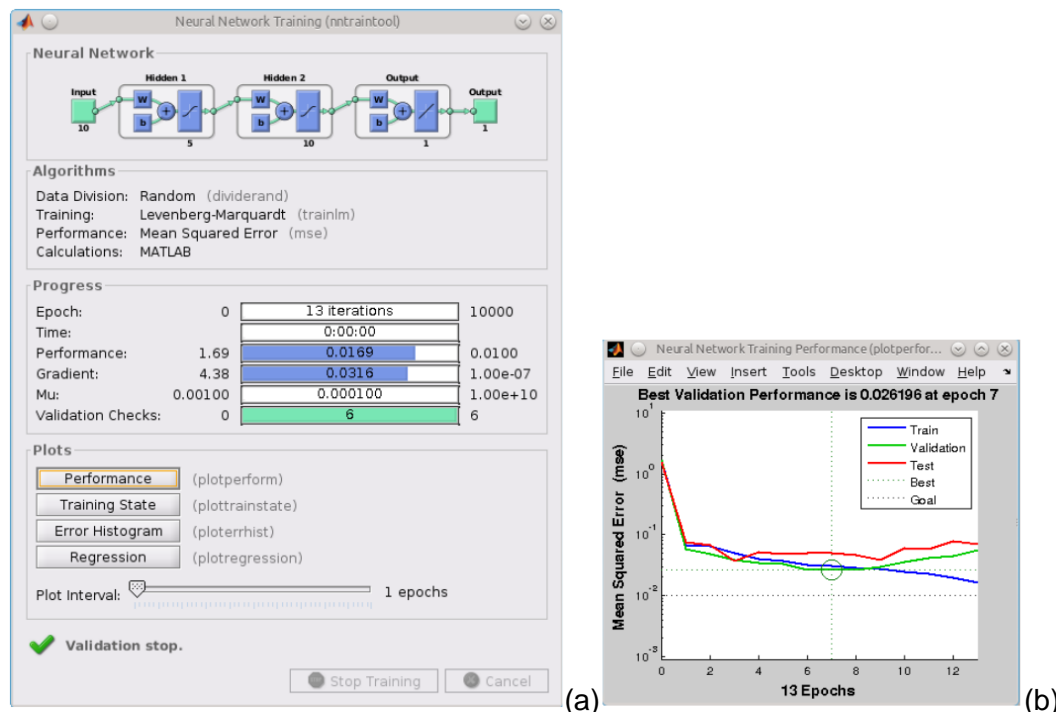


Figura 3. 3: (a) Parámetros de configuración (criterio 1) (b) Curva de convergencia (Arq. 5-10-1)

Elaborado por: Autor

- Criterio de parada #2: Parada anticipada

Según el funcionamiento propio de este criterio de parada, el conjunto de datos debe ser dividido en tres subconjuntos: entrenamiento, validación y prueba. Como se analizó en el capítulo de fundamentación teórica, el subconjunto de entrenamiento se usa para calcular el gradiente y actualizar los pesos de la red, el de validación sirve para monitorizar el error durante el entrenamiento de la red y el de prueba es utilizado para comprobar el desempeño del modelo entrenado. Cuando el error de validación aumenta un determinado número de iteraciones, el entrenamiento es interrumpido y

los pesos se retornan al estado en que estaban cuando se alcanzó el mínimo error.

En la figura 3.4 y figura 3.5 se muestran los parámetros de configuración y las curvas de convergencia para las arquitecturas 3-10-1 y 5-10-1 respectivamente.

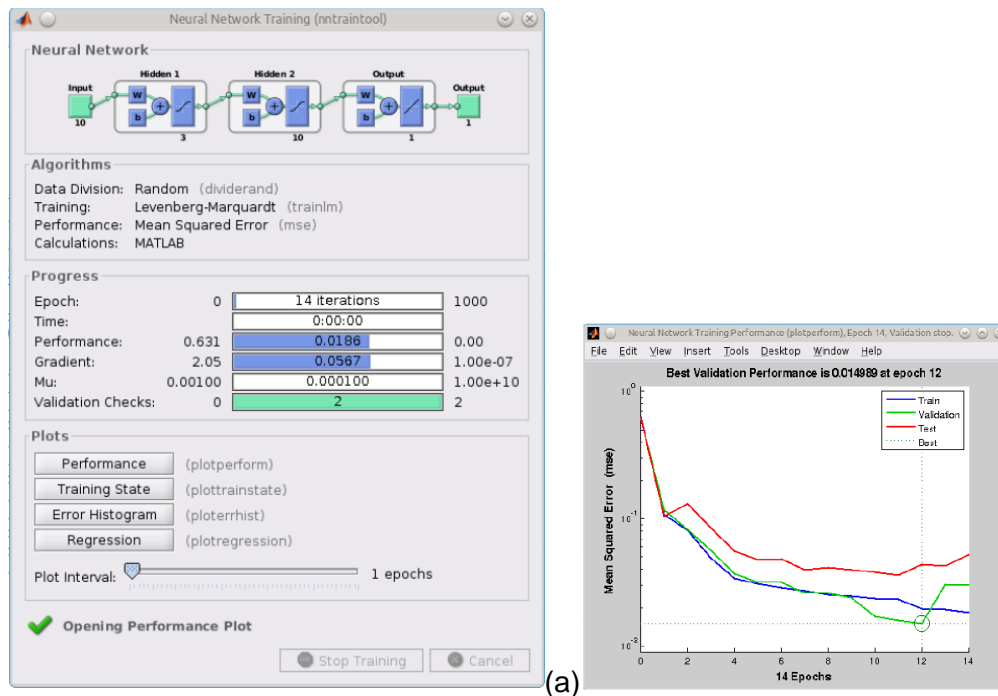


Figura 3. 4: (a) Parámetros de configuración (criterio 2) (b) Curva de convergencia (Arq. 3-10-1)

Elaborado por: Autor

En este caso, los datos se dividieron aleatoriamente en 50% (entrenamiento), 25% (validación) y 25% (prueba) con la función *dividerand*. El número de iteraciones seleccionado para detener el entrenamiento cuando la pérdida en el conjunto de validación comience a crecer es 2: `net.trainParam.max_fail = 2`.

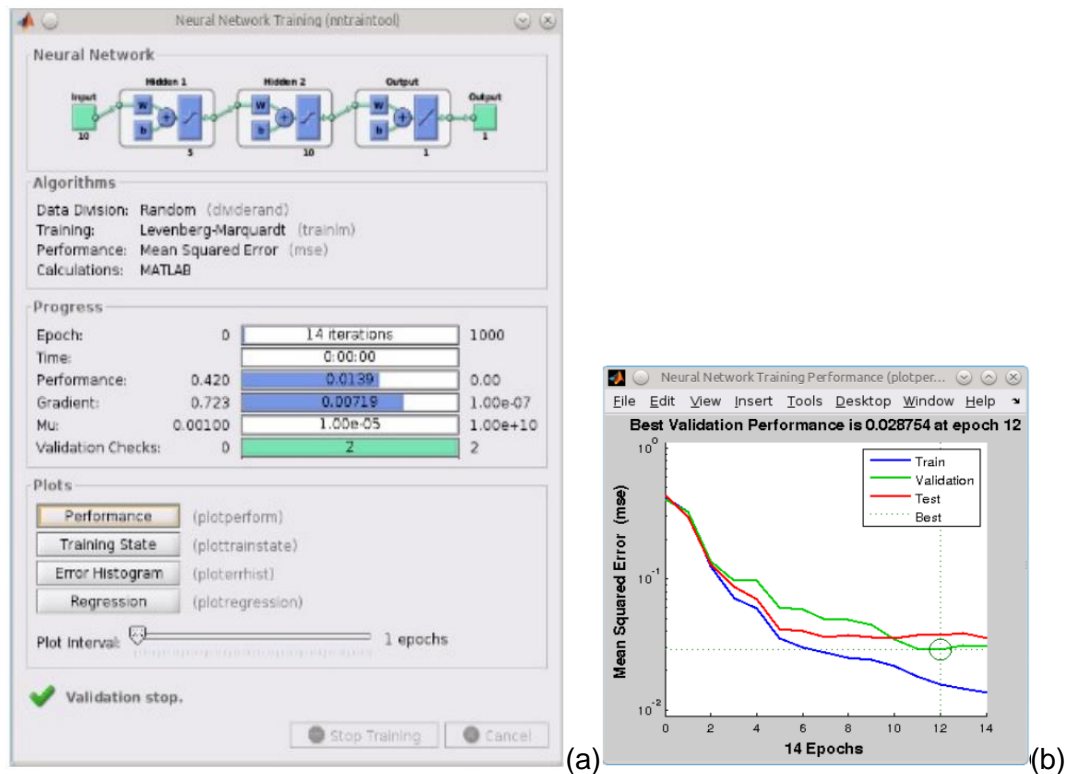


Figura 3. 5: (a) Parámetros de configuración (criterio 2) (b) Curva de convergencia (Arq. 5-10-1)

Elaborado por: Autor

- Criterio de parada #3: Regularización bayesiana

La regularización bayesiana penaliza los pesos de la red, lo cual produce que la respuesta de la red sea más suave y menos propensa al sobreajuste, o sea mejora la generalización. En este caso, al igual que en el criterio 1, los datos se dividieron aleatoriamente mitad y mitad para entrenamiento y prueba con la función *dividerand*.

Generalmente la regulación bayesiana se combina con el algoritmo de optimización de Levenberg – Marquardt, tal y como se ha ejecutado en este caso. En la figura 3.6 y figura .3.7 se muestran los parámetros de configuración y las curvas de convergencia para las arquitecturas 3-10-1 y 5-10-1.

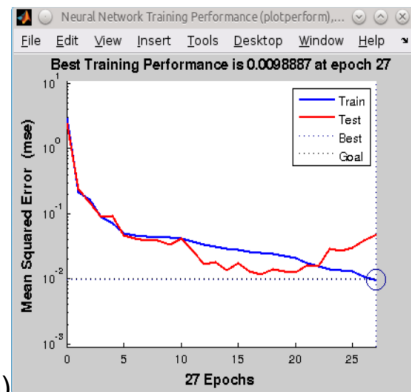
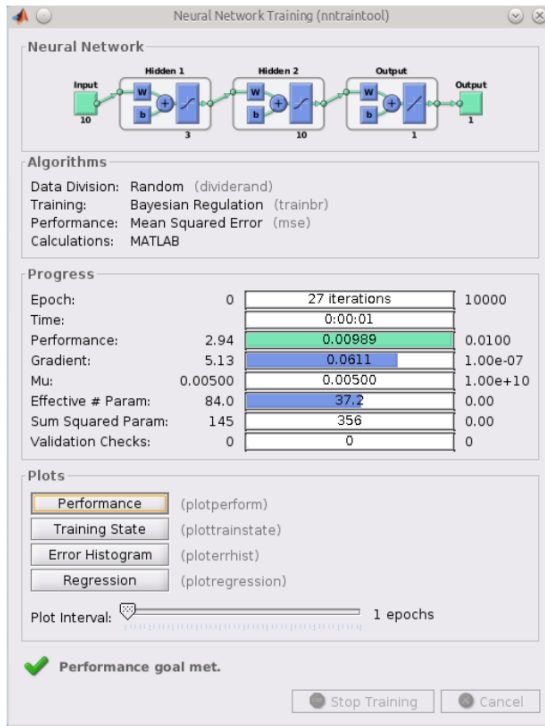


Figura 3. 6: (a) Parámetros de configuración (criterio 3) (b) Curva de convergencia (Arq. 3-10-1)

Elaborado por: Autor

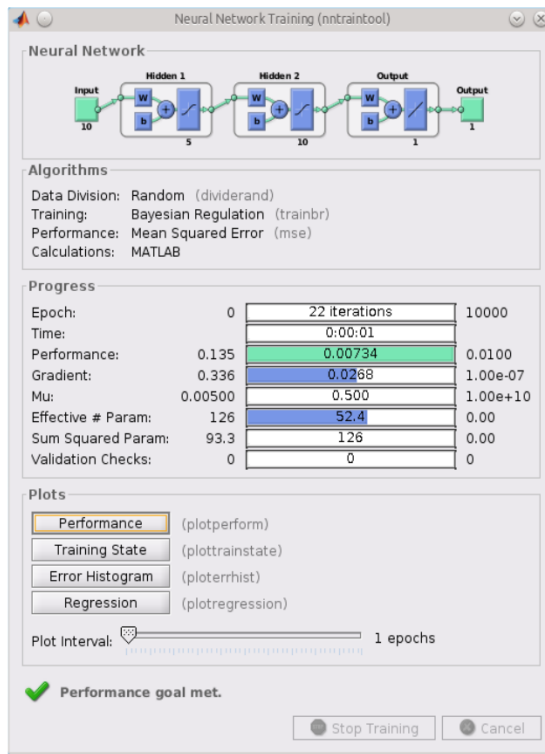
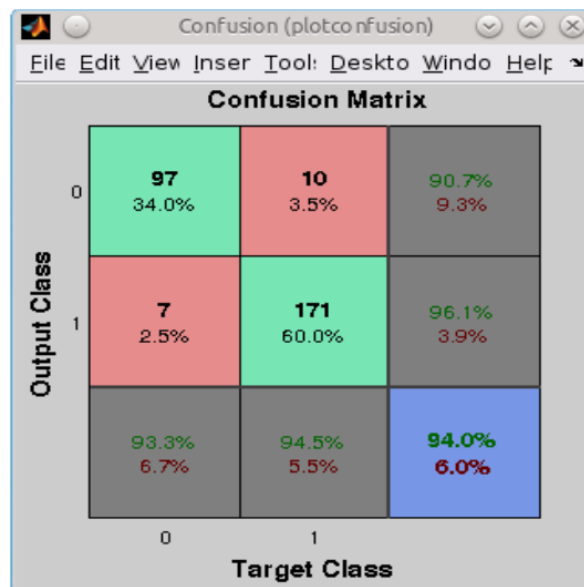


Figura 3. 7: (a) Parámetros de configuración (criterio 3) (b) Curva de convergencia (Arq. 5-10-1)

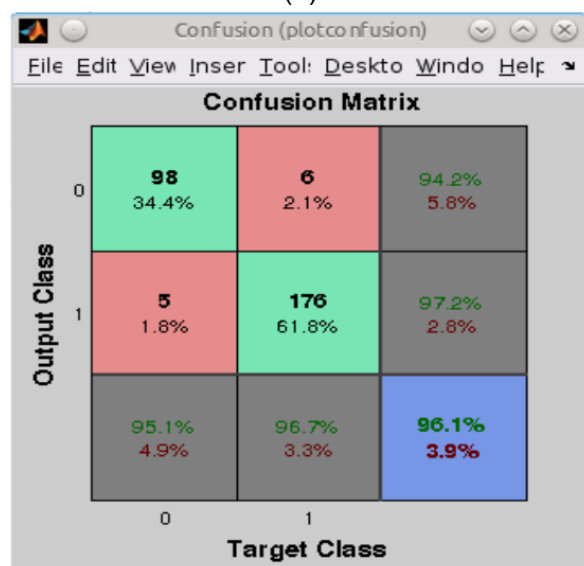
Elaborado por: Autor

3.4. Resultados obtenidos.

En las figuras 3.8, 3.9 y 3.10 se muestran las matrices de confusión para las dos arquitecturas de red y para las tres simulaciones en el conjunto de prueba. La matriz de confusión ofrece una medida efectiva del modelo de clasificación al mostrar el número de clasificaciones correctas e incorrectas contra las clasificaciones predichas para cada clase, sobre un determinado conjunto de datos (Hay, 1988).



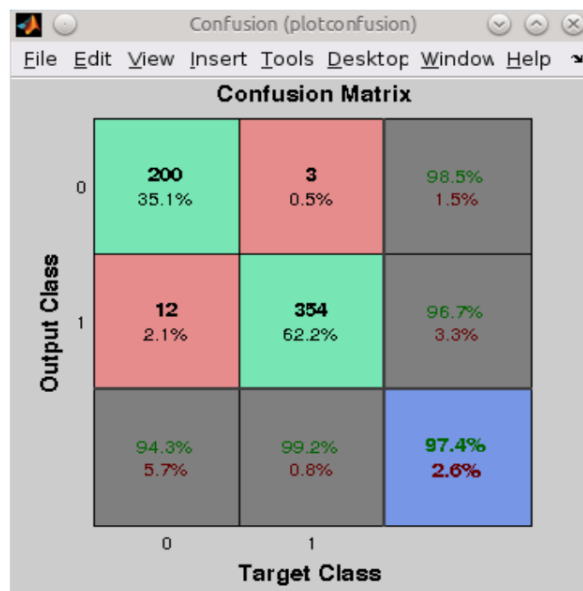
(a)



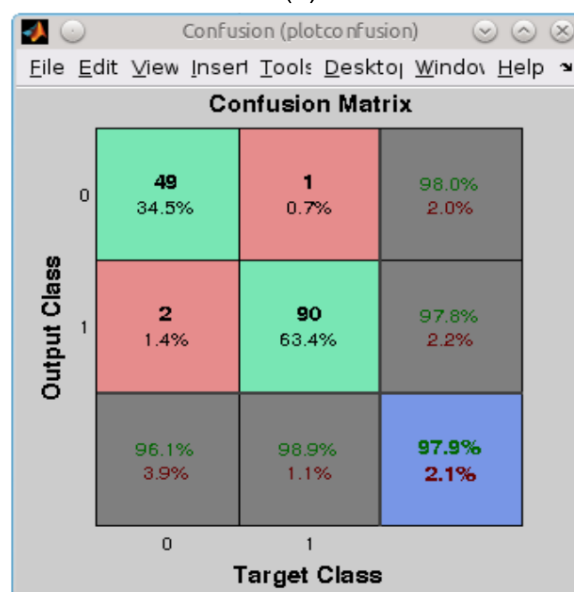
(b)

Figura 3. 8: Matriz de confusión. Criterio: MSE (a) Arq. 3-10-1 (b) Arq. 5-10-1
Elaborado por: Autor

Esta tabla valida el aprendizaje supervisado comparando su base de prueba con la base entrenada e indica la cantidad de aciertos y errores. El número de aciertos, para cada clase, se localiza en la diagonal principal de la matriz. Los demás elementos, para $i \neq j$, representan errores en la clasificación. La matriz de confusión de un clasificador ideal posee todos esos elementos iguales a cero una vez que este no comete ningún error.



(a)

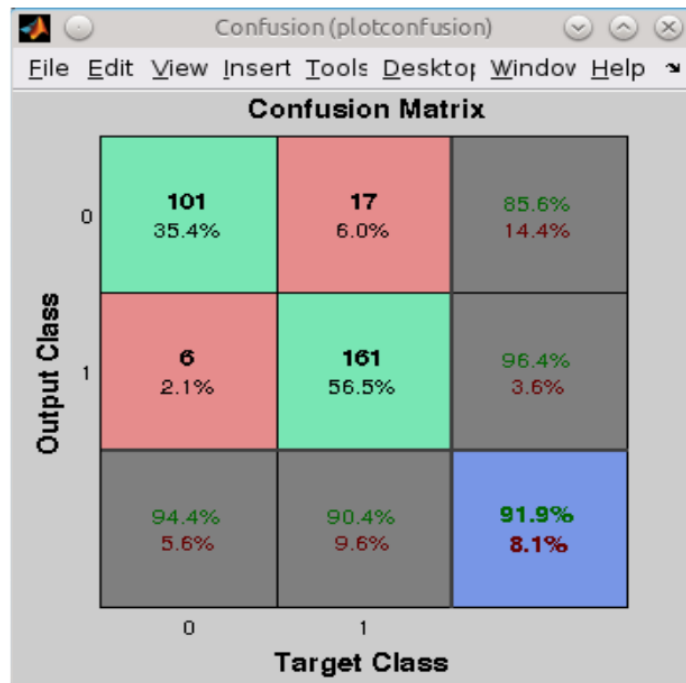


(b)

Figura 3. 9: Matriz de confusión. Criterio: Parada anticipada (a) Arq. 3-10-1 (b) Arq. 5-10-1

Elaborado por: Autor

Con la matriz de confusión se puede obtener la exactitud del experimento, que se define como las clasificaciones correctas dividido por el total de ejemplo.



(a)

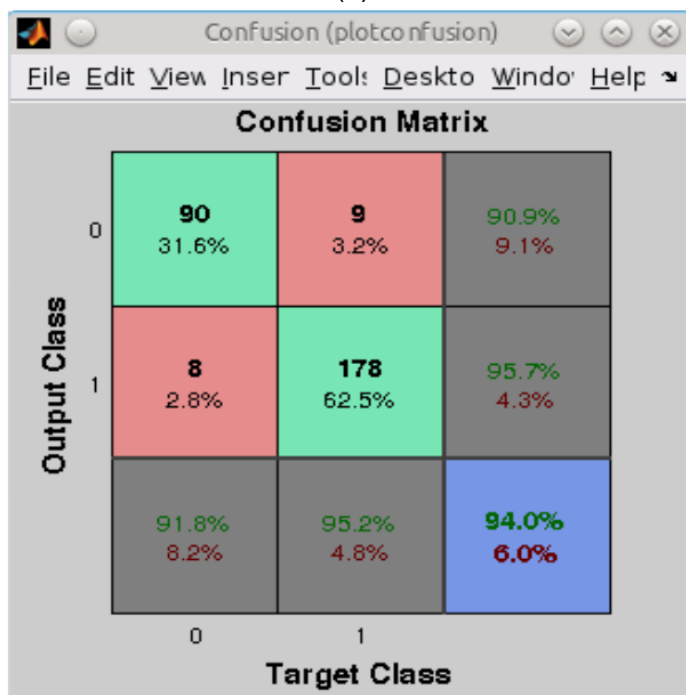


Figura 3. 10: Matriz de confusión. Criterio: Regularización bayesiana (a) Arq. 3-10-1
(b) Arq. 5-10-1

Elaborado por: Autor

El algoritmo fue ejecutado 15 veces, y los valores de exactitud (*accuracy*) obtenidos en las matrices de confusión fueron almacenados. A partir de esos datos fue calculada la exactitud promedio para cada caso. Estos valores son mostrados en la tabla 3.1.

Tabla 3. 1: Valores de exactitud de la clasificación con las dos arquitecturas implementadas y los tres criterios de parada.

# ejecución	Exactitud (criterio 1)		Exactitud (criterio 2)		Exactitud (criterio 3)	
	3-10-1	3-10-5	3-10-1	3-10-5	3-10-1	3-10-5
1	90,90%	92,30%	95,80%	94,40%	88,10%	94,70%
2	96,80%	96,50%	95,80%	96,50%	89,10%	92,30%
3	90,90%	92,30%	91,50%	91,50%	91,60%	95,10%
4	93,70%	95,40%	95,80%	93,00%	94,00%	93,30%
5	93,70%	91,60%	94,40%	95,80%	90,20%	93,70%
6	92,60%	92,30%	94,40%	95,80%	92,30%	90,20%
7	92,60%	87,00%	97,90%	97,90%	93,70%	93,70%
8	94,00%	93,70%	97,90%	95,80%	93,00%	93,70%
9	90,90%	91,60%	95,80%	96,50%	91,60%	91,20%
10	94,00%	95,10%	95,10%	92,30%	92,30%	93,70%
11	93,70%	97,20%	97,20%	98,60%	89,10%	93,70%
12	93,70%	94,40%	94,40%	97,20%	94,00%	89,80%
13	93,70%	93,70%	93,70%	95,10%	91,60%	95,40%
14	93,70%	94,40%	94,40%	94,40%	92,30%	91,60%
15	95,40%	94,40%	94,40%	97,90%	90,90%	92,30%
Promedio	93,35%	93,46%	95,23%	95,51%	91,58%	92,96%

Elaborado por: Autor.

Los resultados obtenidos presentaron desempeños satisfactorios pues todos ellos superan el 90% de acierto en la clasificación. Los mejores valores de exactitud promedio se obtuvieron para las redes entrenadas con el método de aceleración de Levenberg – Marquardt y el criterio de parada

anticipada. Entre todas las redes la de peor resultado fue la arquitectura 3-10-1 usando regularización bayesiana, con una exactitud promedio de 91,58%. De modo general, las arquitecturas 5-10-1 tuvieron mejores desempeños que las 3-10-1 comparadas en relación al mismo criterio de parada.

Conclusiones

En este trabajo se estudiaron los fundamentos teóricos de los sistemas de diagnóstico asistido por computador, su funcionamiento general y las etapas que los componen. Este estudio evidenció la utilidad e importancia de estos sistemas, que contribuyen con la exactitud del diagnóstico y ofrecen una segunda opinión para los especialistas. Además, se abordaron las redes neuronales, las cuales destacan entre los métodos de clasificación usados en los sistemas de diagnóstico asistido por computador. Dichas redes son altamente utilizadas en problemas de clasificación y reconocimiento de patrones en las más variadas áreas de investigación.

Las tecnologías de apoyo al diagnóstico relacionadas con el cáncer de mama resulta un tema de alto impacto, teniendo en cuenta el grave problema de salud que esto representa a nivel mundial. Para la clasificación de biopsias mamarias en benignas o malignas, fueron implementadas dos arquitecturas de redes neuronales artificiales en el Matlab. Cada una fue entrenada con tres criterios de parada. Fue observado que, los mejores valores de exactitud promedio se obtuvieron para las redes entrenadas con el método de aceleración de Levenberg – Marquardt y el criterio de parada anticipada, tanto en la arquitectura 3-10-1 como en la 5-10-1.

Referencias Bibliográficas

- Azevedo-Marquez, P. M. de. (2001). Diagnóstico auxiliado por computador na radiologia. *Radiologia Brasileira*, 34(5), 285–293.
- Caudill, M., & Butler, C. (1992). *Naturally intelligent systems*. MIT press.
- Doan, C. D., & Liong, S.-Y. (2004). Generalization for multilayer neural network: Bayesian regularization or early stopping. In *Proceedings of Asia Pacific Association of Hydrology and Water Resources 2nd Conference*.
- Doi, K. (2007). Computer-Aided Diagnosis in Medical Imaging: Historical Review, Current Status and Future Potential. *Comput Med Imaging Graph.*, 31(4-5), 198–211.
- Falas, T. (1995). Neural networks in empirical accounting research: an alternative to statistical models. *Neural Network World*, 5(4), 419–432.
- Hay, A. M. (1988). The derivation of global estimates from a confusion matrix. *International Journal of Remote Sensing*, 9(8), 1395–1398. <http://doi.org/10.1080/01431168808954945>
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (second edi). Prentice Hall PTR Upper Saddle River, NJ, USA.
- MacKay, D. J. (1991). Bayesian Interpolation. *Neural Computation*, 4, 415–447.
- MacKay, D. J. (1992). A practical Bayesian framework for backpropagation networks. *Neural Computation*, 4(3), 448–472.

- Ministerio de Salud - Ecuador. (2017). Retrieved February 5, 2017, from <http://www.salud.gob.ec/el-ministerio-de-salud-comprometido-en-la-lucha-contra-el-cancer-de-mama/>
- Portal - Cancer Online. (2017). Retrieved February 16, 2017, from http://www.canceronline.cl/index.php?option=com_content&view=article&id=47&Itemid=54
- Ryman-Tubb, N. (1993). A development path to success in neural computing. *Expert Systems Applications*, 9(5), 5–9.
- Shapiro S, Venet W, Strax P, Venet L, R. R. (1982). Ten to fourteen-year effect of screening on breast cancer mortality. *J Natl Cancer Inst*, 69(2), 349–355.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15, 1929–1958.
- Wolberg, W. H., Street, W. N., & Mangasarian, O. L. (1994). Machine learning techniques to diagnose breast cancer from image-processed nuclear features of fine needle aspirates. *Cancer Letters*, 77(2-3), 163–171.



Presidencia
de la República
del Ecuador



Plan Nacional
de Ciencia, Tecnología,
Innovación y Saberes



SENESCYT

Secretaría Nacional de Educación Superior,
Ciencia, Tecnología e Innovación

DECLARACIÓN Y AUTORIZACIÓN

Yo, **ROMERO BERSOZA, MAYRA ALEJANDRA** con C.C: # 0705384121 autor del Trabajo de Titulación: **Implementación de dos arquitecturas de redes neuronales para la clasificación de cáncer de mamas** previo a la obtención del título de **INGENIERA EN TELECOMUNICACIONES** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, 17 de Marzo de 2017

f. _____

Nombre: ROMERO BERSOZA, MAYRA ALEJANDRA

C.C: 0705384121

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TÍTULO Y SUBTÍTULO:	IMPLEMENTACIÓN DE DOS ARQUITECTURAS DE REDES NEURONALES PARA LA CLASIFICACIÓN DE CÁNCER DE MAMAS		
AUTOR(ES)	ROMERO BERSOZA, MAYRA ALEJANDRA		
REVISOR(ES)/TUTOR(ES)	M. Sc. EDWIN FERNANDO PALACIOS MELÉNDEZ		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Facultad de Educación Técnica para el Desarrollo		
CARRERA:	Ingeniería en Telecomunicaciones		
TÍTULO OBTENIDO:	Ingeniero en Telecomunicaciones		
FECHA DE PUBLICACIÓN:	17 de Marzo de 2017	No. DE PÁGINAS:	50
ÁREAS TEMÁTICAS:	Señales y Sistemas, Procesamiento Digital de Señales, Comunicaciones		
PALABRAS CLAVES/ KEYWORDS:	Diagnóstico asistido por computadora, pre-procesamiento, redes neuronales artificiales, cáncer de mama, Overfitting, backpropagation.		
RESUMEN/ABSTRACT (150-250 palabras):			
<p>El cáncer de mama es el cáncer más frecuente en la mujer y representa la primera causa de muerte en países desarrollados y en vías de desarrollo. Su incidencia está aumentando en el mundo debido a la mayor esperanza de vida, el aumento de la urbanización y la adopción de modos de vida occidentales. La identificación del cáncer de mama depende de la interpretación del médico a partir de las informaciones obtenidas de los pacientes a través de exámenes. Las tecnologías para el diagnóstico asistido por computadora dan soporte a los médicos en la interpretación de los contenidos multimedia obtenidos en pruebas a las que se ha sometido el paciente. Las redes neuronales artificiales son fundamentales dentro de estos sistemas. Ellas son un modelo artificial y simplificado del cerebro humano, capaces de adquirir conocimiento a través de la experiencia. En este trabajo son implementadas dos arquitecturas de redes neuronales para la clasificación de biopsias mamarias en malignas y benignas. Las arquitecturas fueron entrenadas con tres criterios de parada cada una, los cuales son: error cuadrático medio, parada anticipada y regularización bayesiana. Los mejores valores de exactitud promedio se obtuvieron para las redes entrenadas con el criterio de parada anticipada.</p>			
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Teléfono: +593-9-59191537	E-mail: alejandra.romerob@gmail.com	
CONTACTO CON LA INSTITUCIÓN: COORDINADOR DEL PROCESO DE UTE	Nombre: Palacios Meléndez Edwin Fernando		
	Teléfono: +593-9-68366762		
	E-mail: edwin.palacios@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			