



Universidad Autónoma de Querétaro
Facultad de Ingeniería
Maestría en Ciencias (Instrumentación y Control)

Determinación en línea del contenido de grasa, lactosa y proteína en leche de ganado bovino mediante redes neuronales

Opción de titulación
Tesis

Que como parte de los requisitos para obtener el Grado de Maestría en Ciencias (Instrumentación y Control)

Presenta:

Ing. José Antonio Licón Gutiérrez

Dirigido por:

Dr. Genaro Martin Soto Zarazúa

Dr. Genaro Martin Soto Zarazúa
Presidente



Firma

Dr. Manuel Toledano Ayala
Secretario



Firma

Dr. Juan Fernando García Trejo
Vocal



Firma

Dr. Luis Miguel Contreras Medina
Suplente

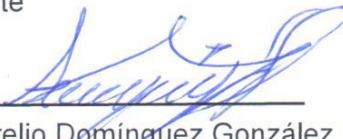


Firma


Dra. Angélica Rosario Jiménez Sánchez
Suplente



Firma



Dr. Aurelio Domínguez González
Director de la Facultad de Ingeniería



Dra. Ma. Guadalupe Flavia Loarca Piña
Director de Investigación y Posgrado

Centro Universitario
Querétaro, Qro.
Junio de 2015

Resumen

Este trabajo presenta la metodología para realizar apropiadamente un diseño experimental, en este caso se realizó el diseño experimental para modelar los parámetros de calidad de la leche; contenido de grasa, lactosa y proteínas. A partir de las variables instrumentales; densidad, conductividad y temperatura. A lo largo de este trabajo se describirá a detalle cómo realizar un modelado de un sistema en concreto por medio de redes neuronales, iniciando por; identificar el tipo de distribución de probabilidad que presenta la población de estudio, diseño experimental, elección de topología de red neuronal y finalmente se presentan los resultados. Cabe señalar que los resultados obtenidos con el modelo basado en redes neuronales fueron de $r_1=0.9505$, $r_2=0.9936$ y $r_3=0.9944$ para el contenido de grasa, lactosa y proteínas respectivamente.

(palabras clave: Calidad de leche, determinación indirecta, redes neuronales, diseño de experimentos y análisis estadísticos)

Summary

This paper presents the methodology for developing a model based on neural networks which can infer indirectly quality parameters of bovine milk (fat, lactose and protein) based on the measurement of density, conductivity and temperature in the sample to be analyzed. To perform this indirect determination is necessary to build a model with which we can relate the measured variables with the variables you want to know. The first step consists in an initial sampling of at least 32 samples; this will serve to estimate the standard deviation and the type of probability distribution that shows the population. The second step is the experimental design; through the design of the experiment we can determine the size of the sampling to will be used for modeling. In the third step is the sampling; at this step a database with which the modeling will be done is performed. In the fourth step the choice of the activation function and type of neural network used is performed; once the above steps are completed, the model based on neural networks is generated. Finally in the fifth step statistical analysis are performed such as analysis of variance, correlation coefficient, Bias coefficient and efficiency coefficient.

(key words: milk quality, indirect determination, neural networks, design of experiments, statistical analysis.)

Agradecimientos

Se agradece al CONACYT, por los recursos brindados para la ejecución de este proyecto, a través de becas y/o estímulos.

Tabla de contenidos

1. INTRODUCCIÓN.....	9
1.1 ANTECEDENTES	10
1.1.1 <i>Artículos científicos.....</i>	<i>10</i>
1.1.2 <i>Patentes relacionadas</i>	<i>13</i>
1.1.3 <i>Productos que están en el mercado</i>	<i>14</i>
1.2 JUSTIFICACIÓN	16
1.3 PLANTEAMIENTO DEL PROBLEMA	17
1.4 HIPÓTESIS Y OBJETIVOS.....	18
1.4.1 <i>Hipótesis</i>	<i>19</i>
1.4.2 <i>Objetivo general</i>	<i>19</i>
1.4.3 <i>Objetivos particulares.....</i>	<i>19</i>
2. MARCO TEÓRICO.....	20
2.1 MUESTREO POR INTERVALOS DE CONFIANZA	20
2.2 INTERVALOS DE CONFIANZA PARA LA MEDIA	24
2.2.1 <i>Varianza poblacional σ^2 conocida</i>	<i>24</i>
2.2.2 <i>Varianza poblacional σ^2 desconocida y $n > 30$</i>	<i>25</i>
2.2.3 <i>Varianza poblacional σ^2 desconocida y $n < 30$</i>	<i>26</i>
2.3 PRUEBA DE NORMALIDAD	27
2.3.1 <i>Prueba de Shapiro y Wilk.....</i>	<i>27</i>
2.3.2 <i>Prueba de Kolmogorov y Smirnov.....</i>	<i>28</i>
2.3.3 <i>Prueba de Lilliefors</i>	<i>28</i>
2.3.4 <i>Prueba de Anderson y Darling</i>	<i>29</i>
2.4 DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA	30
2.5 REDES NEURONALES ARTIFICIALES	31
2.5.1 <i>Elementos básicos que componen una red neuronal</i>	<i>33</i>
2.5.2 <i>Topologías de las redes neuronales.....</i>	<i>38</i>
2.5.3 <i>Mecanismos de aprendizaje para redes neuronales</i>	<i>43</i>
3. METODOLOGÍA	48
3.1 MUESTREO DE RECONOCIMIENTO	48

3.2	PRUEBA DE NORMALIDAD	52
3.3	DETERMINACIÓN DEL TAMAÑO DE LA MUESTRA SEGÚN LOS MÉTODOS TRADICIONALES	56
3.4	DESARROLLO DE LOS MECANISMOS DE APRENDIZAJE.....	56
3.4.1	<i>Mecanismo de aprendizaje función de activación Lineal</i>	57
3.4.2	<i>Mecanismo de aprendizaje función de activación Wavelet</i>	58
3.4.3	<i>Mecanismo de aprendizaje función de activación Sigmoidea</i>	59
3.4.4	<i>Mecanismo de aprendizaje función de activación Tangente hiperbólica</i>	61
3.4.5	<i>Mecanismo de aprendizaje función de activación Gaussiana</i>	62
3.5	MÉTODO DE SELECCIÓN DE LA MEJOR TOPOLOGÍA DE RED NEURONAL	63
3.6	DISEÑO DEL SISTEMA ELECTRÓNICO	67
3.6.1	<i>Medición de densidad</i>	68
3.6.2	<i>Medición de conductividad</i>	70
3.6.3	<i>Medición de temperatura</i>	71
3.6.4	<i>Diseño del sistema mínimo del FPGA</i>	72
3.6.5	<i>Diseño del PCB del sistema completo</i>	73
3.7	IMPLEMENTACIÓN DE LA RED NEURONAL EN FPGA	74
4.	RESULTADOS	75
4.1	RESULTADOS DEL MODELADO EN DISTINTAS TOPOLOGÍAS	75
4.2	RESULTADOS DEL MODELO IMPLEMENTADO EN MATLAB.....	77
4.3	RESULTADOS DEL MODELO IMPLEMENTADO EN FPGA	81
4.4	RESULTADOS DE LA PRUEBA DE VALIDACIÓN DEL MODELO IMPLEMENTADO ¡ERROR! MARCADOR NO DEFINIDO.	
5.	CONCLUSIONES	85
6.	REFERENCIAS.....	85
7.	APÉNDICE.....	¡ERROR! MARCADOR NO DEFINIDO.

Índice de figuras

FIGURA 2.1, Probabilidad dependiendo el numero de desviaciones estándar.	22
FIGURA 2.2, Partes principales de una neurona biológica.	34
FIGURA 2.3, Partes principales de una neurona artificial.	34
FIGURA 2.4, Esquema general de una red neuronal artificial.	38
FIGURA 3.1, Sala de ordeño del campus Amazcala.	49
FIGURA 3.2, Pesador de leche.	49
FIGURA 3.3, LactoscanSP.	50
FIGURA 3.4, Muestras correspondientes al muestro de reconocimiento.	50
FIGURA 3.5, Comparación de distintas topologías de redes neuronales.	67
FIGURA 3.6, Medición de densidad por presión diferencial.	68
FIGURA 3.7, Sensor de presión diferencial.	70
FIGURA 3.8, Sensor de conductividad.	70
FIGURA 3.9, Sensor de temperatura.	71
FIGURA 3.10, Configuración de un dispositivo FPGA.	72
FIGURA 3.11, Top-view del PCB diseñado.	73
FIGURA 3.12, Bottom-view del PCB diseñado.	74

Índice de tablas

TABLA 1.1, Resultados experimentales para el análisis de los principales parámetros de calidad de la leche obtenidos por Pinto et. al. (1998).	11
TABLA 3.1, Resultados del análisis para la muestra de reconocimiento.	51
TABLA 3.2, Prueba de normalidad para la variable contenido de grasa.	53
TABLA 3.3, Prueba de normalidad para la variable contenido de lactosa.	54
TABLA 3.4, Prueba de normalidad para la variable contenido de proteína.	55
TABLA 3.5, Selección de topología para la función de activación Lineal.	64
TABLA 3.6, Selección de topología para la función de activación Wavelet.	64
TABLA 3.7, Selección de topología para la función de activación Sigmoidea.	65
TABLA 3.8, Selección de topología para la función de activación Tangente hiperbólica.	65
TABLA 3.9, Selección de topología para la función de activación Gaussiana.	66

1. Introducción

La leche es considerada como uno de los productos de mayor importancia por su aporte de proteínas de origen animal a la población; por ello para la industria lechera como para el consumidor es importante obtener un producto de buena calidad. El término calidad se puede definir como la suma de las características de los productos y servicios que expresan su capacidad para un uso específico. Relacionado con la producción láctea, se entiende como el conjunto de propiedades físicas y químicas inherentes a la leche y sus derivados que permiten juzgar su valor (sabor, olor, color, valor nutricional y contenido bacteriológico) (Moctezuma et al., 2008). Para el consumidor calidad significa productos de buena presentación, incluyendo olor, color y sabor; para el ganadero significa mayor producción teniendo su hato sano y generando mayores ingresos (Cervantes et al., 2012). En general, la leche de vacas de alta producción, como lo son las razas Holstein o Fresian, está compuesta de agua (87%), grasa (3.8%), proteína (caseína, globulina, lacto-albumina) (3.4%), lactosa (azúcar compuesto por glucosa y galactosa) (4.5%) y otros sólidos totales como minerales (fosfato, sulfato, cloruros), vitaminas (A, D, B1, B2), enzimas (fosfatasa, catalasa, xantinoxidasa, reductasa, peroxidasa, lipasa) y minerales (calcio, sodio, potasio, magnesio) (1.3%) (Calderón et al., 2006). Según Pinto et al. (1998) el contenido de grasa en la leche puede variar de menos 3% a más de 6%, dependiendo de la raza, sistemas de alimentación, clima, periodo de lactancia, mastitis y diferentes formas de manejo de los rebaños. La Cámara Nacional de Industriales de la Leche (CANILEC) (2011a), menciona que la determinación del contenido graso es de gran importancia ya que este parámetro influye en el precio a pagar por litro de leche, permite determinar si una muestra de leche cumple con los valores legales establecidos, estandarizar a los parámetros requeridos para la elaboración de derivados lácteos y tener valores de referencia para la selección genética de los rebaños.

Los métodos utilizados para determinación de grasa en leche y derivados pueden clasificarse dentro de tres grupos (CANILEC, 2011b):

- Métodos volumétricos: utilizan agentes químicos para lograr la ruptura de la emulsión, la separación de la grasa y medir consecutivamente la grasa separada en botellas especiales. A este grupo pertenecen los métodos de Babcock, de Gerber y de la técnica Tesa.
- Métodos gravimétricos: utilizan solventes orgánicos para extraer grasa, que luego de la evaporación de estos, se determina mediante pesada del extracto graso seco. En este grupo se encuentra el método de Roesse-Gottlieb y sus diversas modificaciones como la de Mojonier.
- Métodos instrumentales: fundamentados en la determinación de alguna propiedad de la leche proporcional en algún sentido a su contenido de grasa, por ejemplo la medición de la turbidez en condiciones controladas en instrumentos como el Milkotester, el Lactronic, etc.

1.1 Antecedentes

Durante esta sección se presentaran los antecedentes más destacados sobre el estudio de los parámetros de calidad de la leche, de tal manera que en los capítulos siguientes se puedan entender y evaluar los resultados presentados.

1.1.1 Artículos científicos

El componente lácteo más importante, en el pasado, fue la materia grasa. Su aporte al aroma y sabor de la leche fue ampliamente reconocido, además de su fácil procedimiento de evaluación. El método tradicional es el del Dr. Gerber el cual se practica en la industria lechera desde 1895, fecha en que el Dr. Suizo Nicolás Gerber (1850-1914), propusiera su método denominado "Procedimiento universal para la determinación de la grasa en la leche y los diversos productos lácteos" (Pinto et. al., 1998).

La Federación Internacional de lechería (1983) reconoce el alto valor biológico de las proteínas de la leche, lo cual ya había sido destacado por especialistas en nutrición humana. Provocando con esto cambios en las demandas del consumidor, llegando a tener en la actualidad el contenido de proteínas un valor igual o superior al contenido de grasa (FIL-IDF, 1995).

Pinto et. al. (1998) realizan un experimento en el cual recolectan muestras de leche mensualmente y durante un año en silos de 16 plantas lecheras. Estas fueron analizadas para evaluar sus componentes principales los resultados de este estudio son mostrados en la tabla 1.1.

TABLA 1.1, Resultados experimentales para el análisis de los principales parámetros de calidad de la leche obtenidos por Pinto et. al. (1998).

Parámetro	Promedio (%)	Rango (%)
Grasa	3.53	2.70 - 4.33
Proteína bruta	3.51	3.20 - 3.95
Caseína	2.82	2.03 - 3.33
Proteína del suero	0.68	0.54 - 0.92
Lactosa	4.91	4.26 - 5.57
Sólidos totales	11.94	10.31 - 12.86
Sólidos no grasos	8.42	7.37 - 9.07

Bruckmaier (2001) inicia el tema de los factores que contribuyen a la remoción de leche (ordeño automático) sin embargo no habla sobre la calidad de esta.

Lollivier et. al. (2002) publica que el factor más importante que influencia la remoción de leche es la oxitocina y es crucial para obtener la leche alveolar que es rica en sólidos totales.

Zotte et al., (2006) emplearon un sensor de campo evanescente adelgazado (“taper”) una fibra óptica para la detección del porcentaje de grasa presente en leche líquida. Sumergiendo el taper en muestras de leche durante un tiempo determinado, la potencia óptica disminuye con el porcentaje de grasa presente en la leche, es decir, al haber un número de micelas que interactúan con la componente evanescente de los modos excitados en la cintura del taper, se aumentan las pérdidas de energía óptica.

Pizarro et al., (2007) desarrollaron un sistema de detección en línea basado en espectroscopia de infrarrojo cercano el cual fue utilizado para medir en línea grasa, proteína, lactosa, Cuenta de Células Somáticas (CCS) y nitrógeno ureico en la leche (Milk Urea Nitrogen MUN) durante el proceso de ordeña en un sistema de ordeño automático (AMS por sus siglas en inglés). Este sistema tenía como finalidad proveer a los productores de información sobre la calidad de la leche y la condición fisiológica de cada vaca, proporcionando control de retroalimentación para optimizar la gestión de las granjas lecheras.

Kawamura et. al. (2007) y Yong et. al. (2008) utilizan espectroscopia de infrarrojo cercano para monitorear en línea la calidad de la leche, pero se obtienen resultados distintos en cada experimento.

La Cámara Nacional de Industriales de la Leche Menciona que la determinación del contenido graso es de gran importancia para definir el precio de venta, la fabricación de derivados lácteos y la selección de especímenes para la reproducción (CANILEC, 2011b).

Actualmente el análisis de contenido de grasa en leche se realiza de diferentes formas dependiendo de las posibilidades económicas, las leyes de cada país y las necesidades de los productores.

En México la Norma Oficial Mexicana NOM-155-SCFI-2003, determina que los análisis de calidad se pueden realizar mediante pruebas de laboratorio, instrumentos portátiles o bien los productores se pueden afiliar a la Asociación Holstein de México A.C. En cuanto a pruebas de laboratorio e instrumentos portátiles, estos realizan el análisis tomando muestras de la leche que se desea evaluar y en muchos casos el costo de los mismos es elevado por lo que muchos productores dependen de los análisis que les realizan las empresas a las que les venden sus productos.

1.1.2 Patentes relacionadas

A continuación se presenta un listado de las patentes y solicitudes de patente que están relacionadas con el presente tema de estudio:

En la patente (WO2010142301) de Estonia, la invención comprende un sistema rápido y no invasivo de determinación molecular de oxígeno en la leche ordeñada. El proceso consiste en la medición mediante un sensor de fibra óptica, un dispositivo de medición de amperaje y potencia para la determinación de los niveles de oxígeno, un dispositivo para la adquisición y procesamiento y un dispositivo que genera señales automáticas para la eliminación de la leche que está por debajo de los estándares permitidos que da como resultado la obtención de leche de mejor calidad.

La patente de origen chino (CN102576008) comprende una combinación de métodos para la detección de mastitis y el control de la calidad de la leche. El método consiste en un sensor y un método para la detección de la calidad de la leche. Dicha invención es muy similar a la mencionada anteriormente, puesto que también hace uso de la medición de oxígeno para la determinación de la calidad de la leche y su consecuente eliminación.

La patente (WO2011146001) consiste en una unidad de conteo de células que determina la calidad en términos de la concentración somática celular en la leche. Dicha unidad de conteo celular incluye una cavidad, un almacén temporal de almacenamiento con la intención de su registro mediante imagen y captura de datos. Para cada muestra recogida se capturan dos imágenes, que por medio de un sistema de procesamiento establece una correlación basada en las imágenes de la primera y segunda imagen buscando la presencia de gas de las células somáticas de la muestra de leche. En base a la comparación determina la calidad de la leche.

Otra patente, de origen chino (CN102590460) consiste en la determinación de la calidad de la leche aplicando los siguientes pasos: determinación del contenido de grasa, contenido de proteínas, contenido de fungimicina, colonia de bacterias total y contenido de impurezas. Esta determinación es llevada a cabo mediante un método de normalización de red neuronal, que se ve representada mediante un mapa de radar de acuerdo a los datos recibidos, llevando análisis en conjunto para después transformar el mapa de radar en un mapa de pirámide que lleva a cabo el graduado de la leche mediante procesos matemáticos, como la normalización invertida en el mapa de pirámide. Este proceso permite la evaluación concisa y clara de la calidad de la leche en una imagen en 3D, que es de gran utilidad para los consumidores, empresas y granjeros.

Cabe mencionar que de las invenciones encontradas, solo una, comprende los métodos o sistemas para la realización del monitoreo de forma remota, lo que presenta una importante área de innovación al incorporar los sistemas de determinación de la calidad de la leche. Con la información mostrada anteriormente se puede reconocer que la detección de la calidad de leche está siendo abordada desde diferentes perspectivas de la técnica, aunque desafortunadamente el origen de estas patentes principalmente proviene de otros países, lo que permite entender que la generación de patentes que incorporen los descubrimientos científicos y de la técnica pero desde la perspectiva nacional, podría ser un área de oportunidad de grandes dimensiones.

1.1.3 Productos que están en el mercado

En el mercado de la determinación de parámetros de calidad de la leche existen en la actualidad distintas marcas con una amplia variedad de productos, siendo la mayoría de estos de origen europeo y estadounidenses. En la actualidad no existe un equipo para la determinación de parámetros de calidad de la leche que sea de origen mexicano, lo que ocasiona gastos importantes de envío e

importación, en la mayoría de los productos existentes se usan técnicas de ultrasonido o espectroscopia de infrarrojo cercano, sin embargo, todos los productos existentes trabajan fuera de línea y necesitan aproximadamente de 30 a 90 segundos para realizar la medición, es decir, para poder obtener los parámetros de calidad de la leche es necesario obtener una muestra de leche y trasladarla al laboratorio donde se encuentra el equipo de medición. Esto genera demoras en la medición además de que se deteriora la muestra al entrar en contacto con el medio ambiente y esto ocasiona errores en la medición, por otro lado ninguno de los equipos comerciales cuenta con una base de datos que permita al productor tomar decisiones rápidas y correctas ya sea para mejorar la producción o para mantener la salud de los animales, debido a que presentar variaciones en los parámetros de calidad de la leche puede ser indicador de mala higiene al preparar a las vacas antes y después del proceso de ordeña así como deficiencias en alimentación. A continuación se describen cuatro productos de cuatro marcas diferentes:

MA250 Milk Analyzer: Este sistema de la marca AgriLac de origen estadounidense tiene la capacidad de medir contenido de grasa, sólidos no grasos, densidad, proteínas, lactosa, agua agregada, temperatura, punto de congelamiento y sólidos totales. Este producto está basado en el análisis por medio de ultrasonido y puede realizar 70 muestras por hora en modo continuo. El precio de venta es de mil doscientos cuarenta y seis dólares americanos.

MASTER PRO: Este sistema de la marca MILKOTESTER de origen búlgaro tiene la capacidad de medir contenido de grasa, sólidos no grasos, densidad, proteínas, lactosa, agua agregada, temperatura, punto de congelamiento y sólidos totales. Este producto necesita una muestra de 20ml para operar, el tiempo de análisis es de 60 segundos, este producto está basado en el análisis por medio de ultrasonido. El precio de venta es de mil cuatrocientos dólares americanos.

Lactoscan LA: Este sistema de la marca Lactoscan de origen ruso tiene la capacidad de medir contenido de grasa, sólidos no grasos, densidad, proteínas,

lactosa, agua agregada, temperatura, punto de congelamiento y sólidos totales. El tiempo de operación de este equipo es de 50 segundos y está basado en el análisis por ultrasonido. El precio de venta es de mil trescientos dólares americanos.

MilkoScan FT1: Este sistema de la marca FOSS de origen danés tiene la capacidad de medir contenido de grasa, proteínas, lactosa, sólidos totales, sólidos no grasos, ácidos cítricos, punto de congelamiento, ácidos libres de grasas, densidad y ácidos totales. El tiempo de análisis de este equipo es de 30 segundos y está basado en tecnología de espectroscopia de infrarrojo cercano. El precio de venta es de mil setecientos cincuenta dólares americanos.

1.2 Justificación

La grasa es uno de los parámetros más importantes que se analizan para determinar la calidad de la leche y es clave para definir el precio de la leche fresca y cualquier otro producto final derivado de la misma.

En la industria láctea el color es un indicativo de la riqueza en grasa, ya que el color de la leche depende del tamaño de los glóbulos de grasa, generando un efecto de dispersión de la luz, pues es gracias a su contenido de carotenos que la leche de vaca tiene una tonalidad amarillenta.

El creciente interés en sistemas de pago hacia los productores por parte de la industria láctea, siendo esta última quien establece el precio de la leche con base en el contenido de proteína, grasa y otros índices de calidad (Moctezuma et al., 2008), ha provocado un incremento en la demanda de evaluación en línea de la calidad de la leche al estimular a la industria a buscar soluciones prácticas.

Los métodos actuales de producción láctea implican recolección a granel y almacén prolongado, por lo que es importante mantener un monitoreo constante del contenido de grasa en la producción. Además el uso de la tecnología para

recolectar información en una base de datos automática permite a los productores tener información precisa y vezas acerca de su hato y poder visualizarlos en cualquier momento, permitiendo tomar decisiones correctas ya sea para mejorar la producción o para mantener la salud de los animales, debido a que presentar variaciones en este parámetro puede ser indicador de mala higiene al preparar a las vacas antes y después del proceso de ordeña así como deficiencias en alimentación.

Vázquez-Valencia y Aguilar-Benítez (2010) en un estudio realizado en los Altos de Jalisco concluyeron que se requiere de laboratorios independientes que analicen las muestras de leche para evitar acciones unilaterales por parte de empresas acopiadoras y transformadoras de la materia.

Es por ello que se desarrollara un método para estimar el contenido de grasa, proteína y lactosa en leche bovina. Basado en redes neuronales ya que estas son especialmente útiles para modelar sistemas dinámicos no lineales. Para determinar en línea el contenido de grasa, proteína y lactosa en leche bovina. Será accesible económicamente para los productores, además este sistema contara con una base de datos y una interfaz de usuario permitiendo tener acceso directo y de esta manera tomar decisiones rápidamente.

1.3 Planteamiento del problema

Algunos autores (Castro et. al., 2001) consideran la exigencia sobre los ganaderos para que entreguen la leche mas inocua y de mejor calidad como uno de los principales factores que impulsan la modernización de los sistemas.

En el caso de sistemas familiares como lo menciona Zorrilla et. al. (1997) se presenta la falta de recursos económicos, así que prefieren almacenar su leche en termos colectivos donde les realizan análisis de calidad de la composición de la leche, evaluando factores diariamente como temperatura (°C), volumen (l), acidez

(%), crioscopia (°C), reductasa (minutos), densidad (g/l), grasa (%); así como también la inocuidad (ausencia de sustancias contaminantes en los alimentos) al tomar como parámetros la presencia de antibióticos y/o sedimentos y conteos de células somáticas. Si el producto no cumple con los valores reglamentarios no se acepta el producto provocando que el productor, al no contar con un mercado de venta, tenga que desechar su producción del día, perdiendo ganancias económicas importantes.

Un problema importante con el que se enfrentan los productores lácteos es que las empresas agroindustriales han seguido como estrategia amenazar al productor con dejar de comprarle su producto si no cumple con los requisitos mínimos de calidad.

Castro et. al. (2001) en un estudio realizado en los Altos de Jalisco observaron que por la naturaleza perecedera de la leche para los productores es muy importante contar con un mercado seguro, actualmente no ocurre en la mayoría de los casos porque cuando la leche llega a la industria son ellos los que realizan análisis de calidad al producto que están recibiendo y utilizan esta circunstancia para aplicar castigos injustificados o para rechazar la leche en épocas del año cuando reciben mayor cantidad de la que realmente pueden procesar.

Los métodos tradicionales para determinar contenido de grasa en leche se basan en análisis de muestras en laboratorios especializados o, en algunos casos si se cuenta con recursos económicos, la compra de instrumentos que permitan su análisis ya sea con muestras de leche o en el mejor de los casos que se pueda implementar en línea, la mayoría de estos instrumentos trabajan con tecnología NIR y los costos son elevados.

1.4 Hipótesis y objetivos

1.4.1 Hipótesis

Por medio del uso de redes neuronales en el modelado de sistemas es posible reducir el número de muestras necesarias para realizar un modelado con un nivel de confianza de 95% en los parámetros de calidad de la leche (contenido de grasa, proteína y lactosa). Esto comparado contra el número de muestras teórico que se obtiene por medio del método de "Determinación del tamaño de la muestra por intervalos de confianza".

1.4.2 Objetivo general

Demostrar que por medio del uso de redes neuronales en el modelado de sistemas permite reducir el número de muestras utilizado durante el modelado comparado con el métodos tradicionales. Por medio de un muestreo de reconocimiento se obtendrán los valores que son utilizados por los métodos tradicionales para determinar el tamaño de la muestra por intervalos de confianza, este muestreo preliminar será utilizado para entrenar una red neuronal. Finalmente se somete a la red neuronal a pruebas de validación para valorar si la hipótesis es verdadera o falsa.

1.4.3 Objetivos particulares

1. Diseñar un sistema capaz de medir en línea las variables: temperatura, conductividad, densidad. Que servirán como entradas para la red neuronal.
2. Obtener información numérica de la respuesta del sistema que se desea modelar para un determinado rango de los valores de entrada.

3. Realizar una optimización en cuanto a la topología de la red neuronal que más se adapta al sistema a modelar.

2. Marco teórico

Gorgas et al. (2011) describen en su trabajo que generalmente, una estimación puntual no proporciona un valor exacto del parámetro poblacional a determinar. Es más, en la mayoría de los casos, no tendremos información sobre la precisión de tal estimación, de forma que su valor único no nos informa sobre la probabilidad de que se encuentre cerca o lejos del valor verdadero. En la práctica, interesa no solamente dar una estimación, sino precisar la incertidumbre de dicha estimación. Esto se consigue mediante la estimación por intervalos de confianza.

2.1 Muestreo por intervalos de confianza

Este método está basado en calcular un intervalo sobre el que podamos establecer que, con cierta probabilidad, está contenido el parámetro poblacional desconocido. De esta manera, en vez de calcular un único estimador, se determinan dos estimadores que serán los límites inferior (L1) y superior (L2) (o límites de confianza) de un intervalo de confianza $I = [L1, L2]$. A esta pareja de valores se le llama estimador por intervalo. Estos límites de confianza serán estadísticos que variarán de muestra a muestra, de forma que podrá considerarse al intervalo como una variable aleatoria bidimensional. Efectivamente, los límites del intervalo serán función de los valores que toma la variable aleatoria en el muestreo.

$$L_1 = f_1(X_1, X_2, \dots, X_n) \quad (1)$$

$$L_2 = f_2(X_1, X_2, \dots, X_n) \quad (2)$$

Al valor concreto que toma el intervalo aleatorio en una muestra en particular se le llama estimación por intervalo. Al ser el estimador por intervalo una variable aleatoria, podrá decirse que existe una cierta probabilidad de que el intervalo aleatorio cubra el verdadero valor del parámetro poblacional β . Es decir

$$P(L_1 < \beta < L_2) = 1 - \alpha \quad (3)$$

Donde, por definición, a $1 - \alpha$ se le llama nivel de confianza y al intervalo $[L_1, L_2]$ se le denomina intervalo de confianza del $(1 - \alpha)100\%$.

Nótese que, una vez tomada una muestra en particular, no tiene sentido decir que β estará dentro del intervalo con una cierta probabilidad, puesto que estará o no estará. La forma correcta de expresar esto es diciendo que $1 - \alpha$ es la probabilidad de seleccionar una muestra concreta que conduzca a un intervalo que contenga al parámetro poblacional. En otras palabras, el $100(1 - \alpha)\%$ de los intervalos correspondientes a todas las muestras posibles del mismo tamaño contienen a β y el $100\alpha\%$ no lo contienen.

Evidentemente, al aumentar el tamaño de la muestra ha de aumentar la precisión con que se conoce el parámetro poblacional, y por tanto, para un nivel de confianza fijo, el intervalo de confianza ha de hacerse más pequeño. Es decir, la longitud del intervalo de confianza indica la precisión de la estimación.

Supongamos que para realizar la estimación por intervalos de confianza de un parámetro poblacional se calcula el estadístico B . Este estadístico tendrá una distribución muestral asociada, con media μ_B y desviación típica σ_B . Supongamos que la distribución muestral de B es aproximadamente normal. En este caso, usando las propiedades de la curva normal, podemos establecer las siguientes probabilidades

$$P(\mu_B - \sigma_B < B < \mu_B + \sigma_B) = 0.6827 \quad (4)$$

$$P(\mu_B - 2\sigma_B < B < \mu_B + 2\sigma_B) = 0.9544 \quad (5)$$

$$P(\mu_B - 3\sigma_B < B < \mu_B + 3\sigma_B) = 0.9973 \quad (6)$$

Las ecuaciones 4,5 y 6 pueden ser representadas gráficamente por la figura 2.1 en la cual se presenta el área bajo una curva con distribución normal.

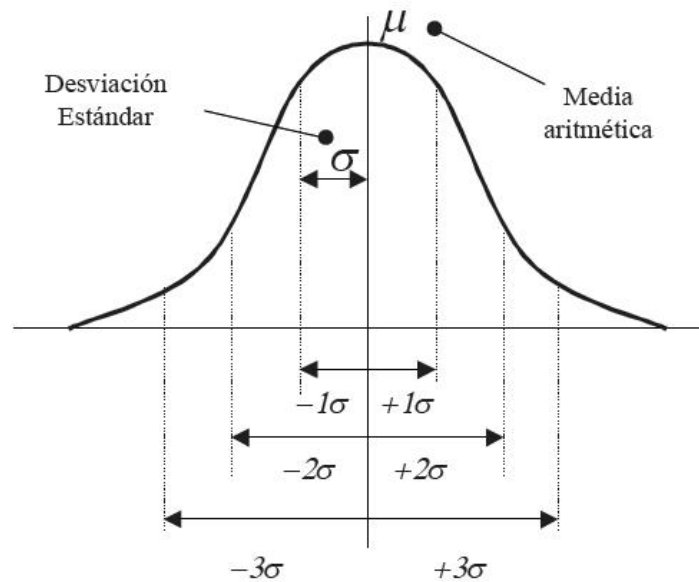


FIGURA 2.1, Probabilidad dependiendo el numero de desviaciones estándar.

Es fácil ver que las ecuaciones 4, 5 y 6 pueden ser representadas por

$$P(B - \sigma_B < \mu_B < B + \sigma_B) = 0.6827 \quad (7)$$

$$P(B - 2\sigma_B < \mu_B < B + 2\sigma_B) = 0.9544 \quad (8)$$

$$P(B - 3\sigma_B < \mu_B < B + 3\sigma_B) = 0.9973 \quad (9)$$

Si B es insesgado, es decir si μ_B coincide con el parámetro poblacional β a determinar, las expresiones anteriores proporcionan intervalos de confianza del

68.27%, 95.44% y 99.73% respectivamente para dicho parámetro poblacional. Normalmente, se suele trabajar con niveles de confianza de 0.95 o 0.99. Para conseguir estas probabilidades hay que buscar en la tabla de la distribución normal las abscisas que dejan a su derecha un área igual a $(1-0.95)/2 = 0.05/2 = 0.025$ y $(1-0.99)/2 = 0.01/2 = 0.005$ respectivamente. Estas son aproximadamente $z_{0.025}=1.96$ y $z_{0.005}=2.58$. Por lo tanto, los intervalos de confianza del 95% y 99% serán respectivamente

$$P(B - 1.96\sigma_B < \mu_B < B + 1.96\sigma_B) = 0.95 \quad (10)$$

$$P(B - 2.58\sigma_B < \mu_B < B + 2.58\sigma_B) = 0.99 \quad (11)$$

En general, para un nivel de confianza $1-\alpha$ habrá que buscar las abscisas $z_{\alpha/2}$ de la distribución normal tipificada $N(0, 1)$ que dejan a su derecha un área igual a $\alpha/2$, expresándose entonces el intervalo de confianza del $(1-\alpha)100\%$ como

$$P(B - z_{\alpha/2}\sigma_B < \mu_B < B + z_{\alpha/2}\sigma_B) = 1 - \alpha \quad (12)$$

La expresión anterior es sumamente útil para calcular intervalos de confianza usando estadísticos con distribuciones muestrales normales. Lo único que habrá que hacer será substituir B por el estadístico insesgado correspondiente y μ_B y σ_B por la media y desviación típica de la distribución muestral.

En el caso de que la distribución muestral del estadístico no sea normal, se pueden hacer las modificaciones correspondientes. Así si B siguiera una distribución t de Student con n grados de libertad, el intervalo vendría dado por

$$P(B - t_{\alpha/2,n}\sigma_B < \mu_B < B + t_{\alpha/2,n}\sigma_B) = 1 - \alpha \quad (13)$$

Donde $t_{\alpha/2,n}$ representa el valor de la abscisa de la distribución t con n grados de libertad que deja a su derecha un área igual a $\alpha/2$. Así mismo, se pueden encontrar las expresiones correspondientes para las distribuciones χ^2 y F , introduciendo las abscisas $\chi^2_{\alpha/2,n}$ y $F_{\alpha/2;n1,n2}$.

2.2 Intervalos de confianza para la media

Supongamos en primer lugar que la población de estudio sigue una distribución normal $N(\mu, \sigma)$ y que como estimador puntual de la media poblacional μ se usa la media muestral \bar{X} . Distinguiremos tres casos principales:

2.2.1 Varianza poblacional σ^2 conocida

Está demostrado que si la población es normal, la media muestral sigue una distribución normal con media $\mu_{\bar{X}}=\mu$ y varianza $\sigma^2_{\bar{X}}=\sigma^2/n$. Entonces, aplicando los términos anteriores en la ecuación 12, el intervalo de confianza del $(1-\alpha)100\%$ para la media puede expresarse como

$$\begin{aligned} P(\bar{X} - z_{\alpha/2}\sigma_{\bar{X}} < \mu_{\bar{X}} < \bar{X} + z_{\alpha/2}\sigma_{\bar{X}}) &= 1 - \alpha \Rightarrow \\ P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) &= 1 - \alpha \end{aligned} \quad (14)$$

Al mismo resultado puede llegarse teniendo en cuenta que, en este caso, la variable $Z=(\bar{X} - \mu)/(\sigma/\sqrt{n})$ es una normal tipificada $N(0,1)$.

$$P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha \quad (15)$$

Entonces la ecuación 15 conduce directamente a la ecuación 14. En resumen, el intervalo de confianza de nivel $(1 - \alpha)$ para la media de una distribución normal de varianza conocida es

$$I = \left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right] \quad (16)$$

El resultado anterior es válido para una población infinita o en un muestreo con reemplazamiento. Si el muestreo es sin reemplazamiento en una población finita de tamaño N , habrá que realizar una modificación a la ecuación 16, de forma que el intervalo de confianza es

$$I = \left[\bar{X} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} \right] \quad (17)$$

Nótese que muestras diferentes darán lugar a valores diferentes de \bar{X} y, por tanto, a intervalos diferentes. Sin embargo, la longitud de los intervalos será siempre la misma y dependerá únicamente (para muestras de igual tamaño) del nivel de confianza $(1 - \alpha)$ que se haya fijado (a menor α mayor anchura del intervalo). Evidentemente, no todos los intervalos que se construyan de diferentes muestras contendrán al parámetro μ , aunque sabemos que esto se cumplirá para el $(1 - \alpha)100\%$ de los intervalos posibles.

2.2.2 Varianza poblacional σ^2 desconocida y $n > 30$

En general, la desviación típica σ de la población se desconoce a priori, de forma que, estrictamente, no se puede aplicar la ecuación 16 para calcular el nivel de confianza. Sin embargo, cuando la muestra es grande, la desviación típica muestral S suele ser un estimador muy preciso de σ , de forma que, en primera aproximación, el intervalo de confianza se puede construir sustituyendo σ por S en la ecuación 16, obteniéndose

$$P\left(\bar{X} - z_{\alpha/2} \frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2} \frac{S}{\sqrt{n}}\right) = 1 - \alpha \quad (18)$$

$$I = \left[\bar{X} \pm z_{\alpha/2} \frac{S}{\sqrt{n}} \right] \quad (19)$$

En la práctica, esta aproximación se usa cuando el tamaño de la muestra n es mayor que 30.

2.2.3 Varianza poblacional σ^2 desconocida y $n < 30$

Cuando las muestras son pequeñas la varianza muestral puede variar considerablemente de muestra a muestra, por lo que la aproximación anterior no se considera válida. En estos casos, el intervalo de confianza se puede construir recordando que la variable $T = (\bar{X} - \mu) / (S / \sqrt{n})$ sigue una distribución t de Student con $n-1$ grados de libertad. Por lo tanto, al ser la distribución t también simétrica, se puede expresar que

$$P\left(\bar{X} - t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}\right) = 1 - \alpha \quad (20)$$

De manera que el intervalo de confianza de nivel $(1 - \alpha)$ para la media de una distribución normal de varianza desconocida y muestra pequeña es

$$I = \left[\bar{X} \pm t_{\alpha/2, n-1} \frac{S}{\sqrt{n}} \right] \quad (21)$$

donde $t_{\alpha/2, n-1}$ es la abscisa de la distribución t que deja a su derecha un área igual a $\alpha/2$. Esta expresión será además exacta y podrá utilizarse para calcular el intervalo de confianza para muestras grandes ($n > 30$). Sin embargo, por las propiedades de la distribución t , esta distribución tiende a la normal al aumentar los grados de libertad, por lo que la ecuación 19 es suficientemente buena si n es grande.

2.3 Prueba de normalidad

De acuerdo con Normadiah y Yap (2011) el que la población de estudio presente una distribución de probabilidad del tipo normal es necesario ya que esta es una suposición de muchos procedimientos estadísticos como: prueba t, análisis de regresión lineal, análisis de discriminante y análisis de varianza (ANOVA). Cuando el supuesto de normalidad no se cumple, la interpretación e inferencias pueden no ser validas. Existen tres procedimientos comunes para determinar el tipo de distribución que presenta una observación de tamaño n , estas son: métodos gráficos; histogramas, graficas de caja y graficas Q-Q, métodos numéricos; oblicuidad e índice de curtosis y finalmente los métodos formales; Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors y Anderson-Darling.

A continuación se describirán los métodos formales por medio de los cuales se puede determinar si la distribución de probabilidad que presenta una muestra de tamaño n es de tipo normal o no.

2.3.1 Prueba de Shapiro y Wilk

La prueba de Shapiro y Wilk (1965) fue originalmente diseñada para muestras con un tamaño de $n < 50$. Esta fue la primera prueba capaz de detectar desviaciones en la normalidad ya sea debido a la asimetría, curtosis o ambas. Según Mandes y Pala (2003) esta se ha convertido en una de las pruebas más utilizadas debido a que arroja buenos resultados. Dada una muestra aleatoria ordenada, $y_1 < y_2 < \dots < y_n$, se define la prueba como

$$W = \frac{(\sum_{i=1}^n a_i y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (22)$$

Donde y_i son los valores que toma cada una de las muestras, \bar{y} es la media aritmética de las muestras, $a_i = (a_1, \dots, a_n) = (m^T v^{-1}) / (m^T v^{-1} v^{-1} m)^{1/2}$ y $m = (m_1, \dots, m_n)^T$ son

los valores esperados de una muestra de la distribución normal estándar y finalmente v es la matriz de covarianza de los datos m . El valor de W se encontrara en el rango de $0 \leq W \leq 1$. Para valores pequeños de W justifican el rechazo de normalidad mientras que un valor de uno indica la normalidad de los datos.

2.3.2 Prueba de Kolmogorov y Smirnov

Esta prueba pertenece a la clase suprema de la estadística de función de distribución empírica EDF y está basada en la diferencia vertical entre la muestra a evaluar y la distribución empírica (Conover, 1999). Dada n con datos ordenados, $x_1 < x_2 < \dots < x_n$, Conover (1999) redefinió esta prueba estadística propuesta por Kolmogorov (1933) como

$$T = \text{SUP}_x |F^*(x) - F_n(x)| \quad (23)$$

Donde SUP es la función "el mayor". $F^*(x)$ es la función de la muestra a evaluar mientras que $F_n(x)$ es la función de datos obtenidos al azar de una distribución empírica.

2.3.3 Prueba de Lillefors

Esta prueba es una modificación de la prueba de Kolmogorov-Smirnov. Esta prueba es apropiada en una situación donde los parámetros de la muestra a analizar son completamente conocidos. Sin embargo, algunas veces es difícil especificar los parámetros estadísticos ya que el tipo de distribución es desconocido. En este caso, los parámetros necesitan ser estimados basándose en los datos de la muestra. Cuando se usa la prueba de Kolmogorov-Smirnov es usada en esta situación, el resultado puede ser engañoso por lo que la

probabilidad de error tiende a ser más pequeña que la que aparece en la tabla estándar de la prueba de Kolmogorov (Lilliefors, 1967). Por lo tanto, en esta situación es preferida la prueba de Lilliefors. Dada una muestra de n observaciones, la prueba de Lilliefors está definida como

$$D = \text{MAX}_x |F^*(x) - S_n(x)| \quad (24)$$

Donde $S_n(x)$ es la función de distribución acumulada de la muestra a analizar y $F^*(x)$ es la función de distribución normal acumulada con $\mu = \bar{X}$ como la media aritmética de la muestra, S^2 la varianza de la muestra definida con el denominador $n-1$.

2.3.4 Prueba de Anderson y Darling

Esta prueba difiere de las pruebas anteriores, en el sentido que brinda una mayor contribución a los valores que se encuentran en las colas de la distribución (Farrel y Stewart, 2006). De acuerdo con Arshad et al. (2003) este método es la más poderosa prueba basada en una función de distribución empírica. Está basada en la diferencia cuadrática $(F_n(x) - F^*(x))^2$, Anderson y Darling (1954) definieron su prueba estadística como

$$A^2 = n \int_{-\infty}^{\infty} [F_n(x) - F^*(x)]^2 \psi(F^*(x)) dF^*(x) \quad (25)$$

Donde ψ es una función de peso no negativa que puede ser calculada como $\psi = [F^*(x)(1 - F^*(x))]^{-1}$. Con el fin de hacer el cálculo estadístico con mayor facilidad en la práctica, la siguiente ecuación puede ser aplicada (Arshad et al., 2003)

$$A^2 = -n - \frac{1}{n} \sum_{i=1}^n [(2i-1)(\log F^*(x_i) + \log(1 - F^*(x_{n+1-i})))] \quad (26)$$

Donde $F^*(x_i)$ es la función de distribución acumulada, x_i son los datos ordenados y n es el tamaño de la muestra.

2.4 Determinación del tamaño de la muestra

Hasta ahora siempre se ha supuesto conocido el tamaño de la muestra n . Sin embargo, y fundamentalmente en el diseño de experimentos, en ocasiones el problema principal es la determinación del tamaño muestral requerido para obtener la estimación de los parámetros poblacionales con una determinada precisión. Nótese que una muestra demasiado grande puede traducirse en una pérdida de tiempo y dinero, mientras que, si la muestra es pequeña, no se obtendrá la fiabilidad deseada y el experimento será un fracaso.

La precisión de una estimación por intervalos de confianza vendrá marcada por la longitud del intervalo (en ocasiones, llamado error). Supongamos que tenemos una distribución normal y queremos determinar la media poblacional μ a partir de la media muestral \bar{X} . El intervalo de confianza está dado por la ecuación 19, de manera que la longitud del intervalo se puede expresar como

$$l = 2z_{\alpha/2}(S/\sqrt{n}) \quad (27)$$

Es decir, la longitud del intervalo es inversamente proporcional al tamaño de la muestra y la precisión aumenta, por tanto, al aumentar n . El problema se plantea entonces en cómo calcular el tamaño de la muestra n para estimar la media poblacional con una cierta precisión, es decir, para que la diferencia entre la media poblacional y muestral sea, en valor absoluto y con cierto nivel de confianza $(1 - \alpha)$, menor que un cierto error, denotado por ε

$$P(\bar{X} - \varepsilon < \mu < \bar{X} + \varepsilon) = 1 - \alpha \quad (28)$$

De esta forma, comparando la expresión anterior con la ecuación 14, una vez fijado α puede calcularse n igualando el error ε a la semilongitud del intervalo ($l/2$)

$$\varepsilon = z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \Rightarrow n = z_{\alpha/2}^2 \frac{\sigma^2}{\varepsilon^2} \quad (29)$$

Para poder aplicar la expresión anterior es necesario conocer previamente σ . Si éste no es el caso, en la práctica se toma una muestra piloto pequeña (aunque es deseable que $n > 30$) para poder estimar σ mediante la desviación típica muestral S .

2.5 Redes neuronales artificiales

El hombre se ha caracterizado siempre por su búsqueda constante de nuevas vías para mejorar sus condiciones de vida. Estos esfuerzos le han servido para reducir el trabajo en aquellas operaciones en las que la fuerza juega un papel primordial. Los progresos obtenidos han permitido dirigir estos esfuerzos a otros campos, como por ejemplo, a la construcción de máquinas calculadoras que ayuden a resolver de forma automática y rápida determinadas operaciones que resultan tediosas cuando se realizan a mano.

Estas máquinas permiten implementar fácilmente algoritmos para resolver multitud de problemas que antes resultaban engorrosos de resolver. Sin embargo, se observa una limitación importante: ¿qué ocurre cuando el problema que se quiere resolver no admite un tratamiento algorítmico, como es el caso, por ejemplo, de la clasificación de objetos por rasgos comunes? Este ejemplo demuestra que la construcción de nuevas máquinas más versátiles requiere un enfoque del problema desde otro punto de vista. Los desarrollos actuales de los científicos se dirigen al estudio de las capacidades humanas como una fuente de nuevas ideas para el diseño de las nuevas máquinas. Así, la inteligencia artificial

es un intento por descubrir y describir aspectos de la inteligencia humana que pueden ser simulados mediante máquinas. Esta disciplina se ha desarrollado fuertemente en los últimos años teniendo aplicación en algunos campos como visión artificial, demostración de teoremas, procesamiento de información expresada mediante lenguajes humanos.

Las redes neuronales son más que otra forma de emular ciertas características propias de los humanos, como la capacidad de memorizar y de asociar hechos. Si se examinan con atención aquellos problemas que no pueden expresarse a través de un algoritmo, se observará que todos ellos tienen una característica en común: la experiencia. El hombre es capaz de resolver estas situaciones acudiendo a la experiencia acumulada. Así, parece claro que una forma de aproximarse al problema consista en la construcción de sistemas que sean capaces de reproducir esta característica humana. En definitiva, las redes neuronales no son más que un modelo artificial y simplificado del cerebro humano, que es el ejemplo más perfecto del que disponemos para un sistema que es capaz de adquirir conocimiento a través de la experiencia. Una red neuronal es “un nuevo sistema para el tratamiento de la información, cuya unidad básica de procesamiento está inspirada en la célula fundamental del sistema nervioso humano: la neurona”.

Debido a su constitución y a sus fundamentos, las redes neuronales artificiales presentan un gran número de características semejantes a las del cerebro. Por ejemplo, son capaces de aprender de la experiencia, de generalizar de casos anteriores a nuevos casos, de abstraer características esenciales a partir de entradas que representan información irrelevante, etc. Esto hace que ofrezcan numerosas ventajas y que este tipo de tecnología se esté aplicando en múltiples áreas. Entre las ventajas se incluyen:

- Aprendizaje adaptativo: Capacidad de aprender a realizar tareas basadas en un entrenamiento o en una experiencia inicial.

- Auto-organización: Una red neuronal puede crear su propia organización o representación de la información que recibe mediante una etapa de aprendizaje.
- Tolerancia a fallos: La destrucción parcial de una red conduce a una degradación de su estructura; sin embargo, algunas capacidades de la red se pueden retener, incluso sufriendo un gran daño.
- Operación en tiempo real: Los cálculos neuronales pueden ser realizados en paralelo; para esto se diseñan y fabrican máquinas con hardware especial para obtener esta capacidad.
- Fácil inserción dentro de la tecnología existente: Se pueden obtener chips especializados para redes neuronales que mejoran su capacidad en ciertas tareas. Ello facilitará la integración modular en los sistemas existentes.

2.5.1 Elementos básicos que componen una red neuronal

Antes de describir con mayor detalle una red neuronal, se debe de aprender sobre las neuronas y de cómo ellas son utilizadas por una red neuronal. En la figura 2.2 se describen las partes principales de una neurona biológica, mientras que en la figura 2.3 se muestra una neurona artificial en la cual se contrastan las principales partes de una neurona biológica con las de una neurona artificial.

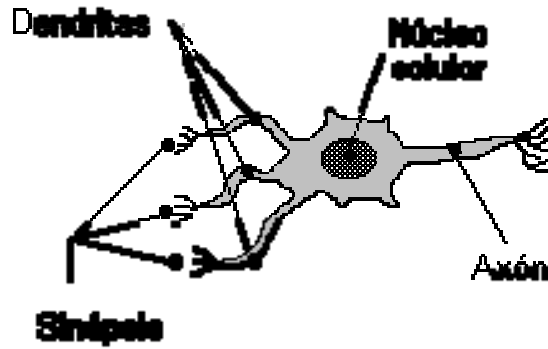


FIGURA 2.2, Partes principales de una neurona biológica.

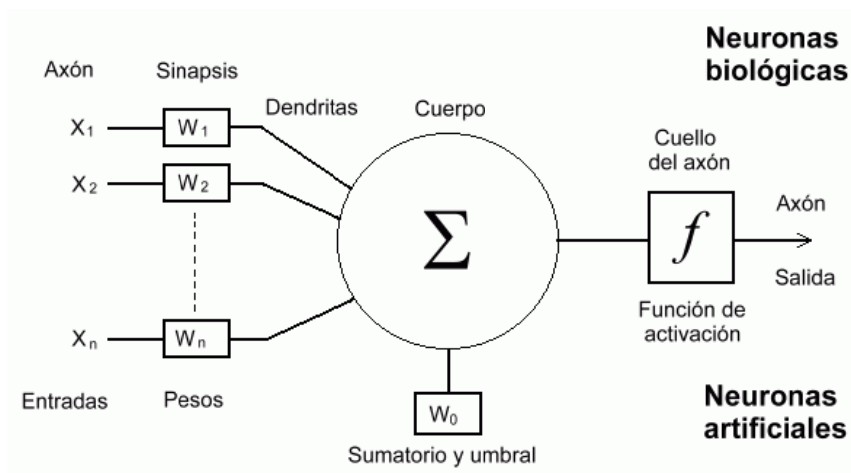


FIGURA 2.3, Partes principales de una neurona artificial.

De las figuras anteriores se puede observar que existen similitudes entre ambas ya que tienen entradas, utilizan pesos de interconexión y generan respuestas en su salida las cuales dependen de la información que se encuentra en los pesos de interconexión.

2.5.1.1 Función de entrada

La neurona trata a muchos valores de entrada como si fueran uno solo; esto recibe el nombre de entrada global. Por lo tanto, ahora nos enfrentamos al problema de cómo se pueden combinar estas entradas X_1, X_2, \dots, X_n con los pesos de interconexión W_1, W_2, \dots, W_n para así generar la entrada global S . Esto se logra a través de la función de entrada, la cual se calcula a partir del vector de entrada, el vector de pesos de interconexión y los operadores matemáticos con los que deseamos relacionarlos. La función de entrada puede describirse como sigue:

$$S = (X_1 * W_1) + (X_2 * W_2) + \dots + (X_n * W_n) \quad (30)$$

Donde n es el número de entradas a la neurona, cabe señalar que el operador sumatoria puede ser cambiado por otro operador como por ejemplo: máximo, mínimo, producto, etc. según sea la aplicación. De la ecuación 30 se observa que los valores de entrada se multiplican por los pesos de interconexión. 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 suficientemente pequeños.

2.5.1.2 *Función de activación*

Ya que 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 calcula el estado de actividad de una neurona; transformando la entrada global en un valor de activación, cuyo rango normalmente va de $[0, 1]$ o $[-1, 1]$. Las funciones de activación más comúnmente utilizadas se detallan a continuación:

- 1) Función Lineal: Los valores de salida obtenidos por medio de esta función de activación serán: $a \cdot S$, cuando el argumento S este comprendido dentro del rango $(-1/a, 1/a)$. Por encima o por debajo de esta zona se fija la salida en 1 o -1, respectivamente. La función matemática se define en la ecuación 31.

$$f(x) = \begin{cases} -1 & x \leq -1/a \\ a \cdot x & -1/a < x < 1/a \\ 1 & x \geq 1/a \end{cases} \quad (31)$$

$a > 0$

- 2) Función Wavelet: Los valores de salida obtenidos por medio de esta función de activación estarán definidos por la ecuación 32, en la cual se ha decidido utilizar $a=1$ y $b=0$ esto para facilitar los cálculos en los temas siguientes.

$$\begin{aligned} \psi_{a,b}(x) &= |a|^{-1/2} \psi((x-b)/a) \\ \psi(t) &= (1-t^2) * e^{(-t^2/2)} \\ \text{Para: } a &= 1, b = 0 \\ \psi_{a,b}(x) &= (1-x^2) * e^{(-x^2/2)} \end{aligned} \quad (32)$$

- 3) Función Sigmoidea: Los valores de salida que proporciona esta función están comprendidos dentro de un rango que va de 0 a 1. En la ecuación 33 se muestra la representación matemática de esta función.

$$f(x) = \frac{1}{1 + e^{-x}} \quad (33)$$

- 4) Función Tangente Hiperbólica: Los valores de salida de la función tangente hiperbólica están comprendidos dentro de un rango que va

de -1 a 1. En la ecuación 34 se muestra la representación matemática de esta función.

$$f(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (34)$$

5) Función Gaussiana: Los valores de esta función están comprendidos dentro de un rango que va de 0 a 1. En la ecuación 35 se muestra la representación matemática de esta función.

$$f(x) = e^{-x^2} \quad (35)$$

2.5.1.3 Función de salida

El último componente que una neurona necesita es la función de salida de la neurona, es por esto que, la función de salida determina que valor se transfiere a las neuronas vinculadas. Si la función de activación está por debajo de un umbral determinado, ninguna salida se pasa a la neurona subsiguiente. Normalmente, no cualquier valor es permitido como una entrada para una neurona, por lo tanto, los valores de salida están comprendidos en el rango $[0, 1]$ o $[-1, 1]$. También pueden ser binarios $\{0, 1\}$ o $\{-1, 1\}$. Dos de las funciones de salida más comunes son:

1) Transparente: Este es el tipo de función más sencillo, tal que la salida es la misma que la entrada. También es llamada función identidad. En la ecuación 36 se muestra la representación matemática de esta función.

$$f(x) = x \quad (36)$$

2) Binaria: Este tipo de función toma un valor de 1 cuando la entrada a la función de salida supera un cierto umbral, de otro modo el valor de la salida es 0. En la ecuación 37 se muestra la representación matemática de esta función. Donde ξ es el umbral.

$$f(x) = \begin{cases} 1 & x \geq \xi \\ 0 & x < \xi \end{cases} \quad (37)$$

En la figura 2.4 se puede observar el esquema general de una red neuronal artificial:

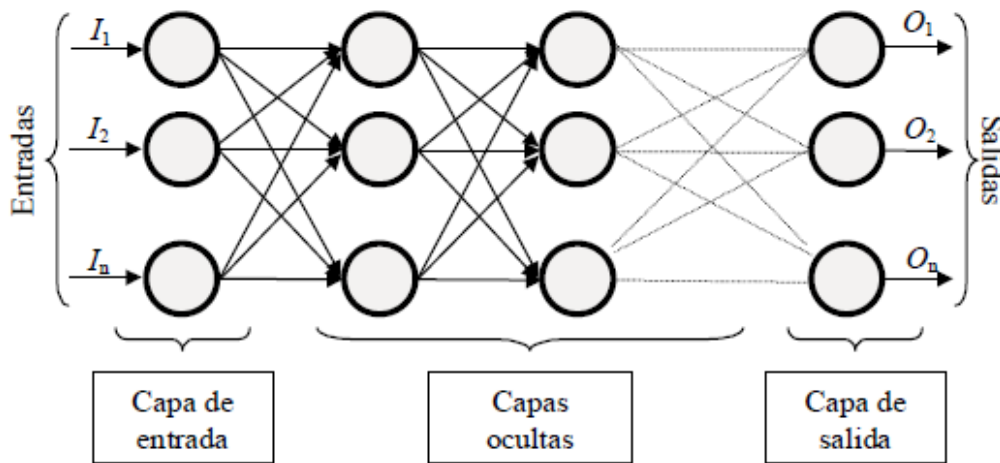


FIGURA 2.4, Esquema general de una red neuronal artificial.

El esquema general está constituido por neuronas interconectadas y arregladas en tres secciones principales. Los datos ingresan por medio de la capa de entrada, pasan a través de la(s) capa(s) oculta(s) y salen por la capa de salida. Cabe mencionar que la capa oculta puede estar constituida por varias capas.

2.5.2 Topologías de las redes neuronales

La topología o arquitectura de una red neuronal consiste en la organización y disposición de las neuronas en la misma, formando capas o agrupaciones de

neuronas más o menos alejadas de la entrada y salida de dicha red. En este sentido, los parámetros fundamentales de la red son: el número de capas, el número de neuronas por capa, el grado de conectividad y el tipo de conexiones entre neuronas.

2.5.2.1 Redes monocapa

En las redes monocapa, se establecen conexiones entre las neuronas que pertenecen a la única capa que constituye la red. Las redes monocapas se utilizan generalmente en tareas relacionadas con lo que se conoce como autoasociación (regenerar información de entrada que se presenta a la red de forma incompleta o distorsionada).

2.5.2.2 Redes multicapa

Las redes multicapas son aquellas que disponen de un conjunto de neuronas agrupadas en varios niveles o capas. En estos casos, una forma para distinguir la capa a la que pertenece una neurona, consistiría en fijarse en el origen de las señales que recibe a la entrada y el destino de la señal de salida. Normalmente, todas las neuronas de una capa reciben señales de entrada desde otra capa anterior (la cual está más cerca a la entrada de la red), y envían señales de salida a una capa posterior (que está más cerca a la salida de la red). A estas conexiones se las denomina conexiones hacia adelante o feedforward.

Sin embargo, en un gran número de estas redes también existe la posibilidad de conectar la salida de las neuronas de capas posteriores a la entrada de capas anteriores; a estas conexiones se las denomina conexiones hacia atrás o feedback.

Estas dos posibilidades permiten distinguir entre dos tipos de redes con múltiples capas: las redes con conexiones hacia adelante o redes feedforward, y las redes que disponen de conexiones tanto hacia adelante como hacia atrás o redes feedforward/feedback.

2.5.2.3 Conexión entre neuronas

La conectividad entre los nodos de una red neuronal está relacionada con la forma en que las salidas de las neuronas están canalizadas para convertirse en entradas de otras neuronas. La señal de salida de un nodo puede ser una entrada de otro elemento de proceso, o incluso ser una entrada de sí mismo (conexión autorrecurrente).

Cuando ninguna salida de las neuronas es entrada de neuronas del mismo nivel o de niveles precedentes, la red se describe como de conexión hacia adelante. Cuando las salidas pueden ser conectadas como entradas de neuronas de niveles previos o del mismo nivel, incluyéndose ellas mismas, la red es de conexión hacia atrás. Las redes de propagación hacia atrás que tienen lazos cerrados son llamadas sistemas recurrentes.

2.5.2.4 Asociaciones entre la información de entrada y salida

Ya se sabe que las redes neuronales son sistemas que almacenan cierta información aprendida. Esta información se registra de forma distribuida en los pesos asociados a las conexiones entre neuronas. Por tanto, puede imaginarse una red como cierto tipo de memoria que almacena datos de forma estable, datos que se grabarán en dicha memoria como consecuencia del aprendizaje de la red y que podrán ser leídos a la salida como respuesta a cierta información de entrada,

comportándose entonces la red como lo que habitualmente se conoce por memoria asociativa: cuando se aplica un estímulo (dato de entrada) la red responde con una salida asociada a dicha información de entrada.

Existen dos formas primarias de realizar esta asociación entre entradas/salidas que se corresponden con la naturaleza de la información almacenada en la red. Una primera sería la denominada heteroasociación, que se refiere al caso en el que la red aprende parejas de datos $[(A_1, B_1), (A_2, B_2), \dots, (A_N, B_N)]$, de tal forma que cuando se presente cierta información de entrada A_i , deberá responder generando la correspondiente salida asociada B_i . La segunda se conoce como autoasociación, donde la red aprende ciertas informaciones A_1, A_2, \dots, A_N ; de tal forma que cuando se le presenta una información de entrada realizará una autocorrelación, respondiendo con uno de los datos almacenados, el más parecido al de entrada.

Estos dos mecanismos de asociación dan lugar a dos tipos de redes neuronales: las redes heteroasociativas y las autoasociativas. Una red heteroasociativa podría considerarse como aquella que computa cierta función, que en la mayoría de los casos no podría expresarse analíticamente, entre un conjunto de entradas y un conjunto de salidas, correspondiendo a cada posible entrada una determinada salida. Por otra parte, una red autoasociativa es una red cuya principal misión es reconstruir una determinada información de entrada que se presente incompleta o distorsionada (le asocia el dato almacenado más parecido).

En realidad estos dos tipos de modelos de redes no son diferentes en principio, porque una red heteroasociativa puede siempre ser reducida a una asociativa mediante la concatenación de una información de entrada y su salida (respuesta) asociada, para obtener la información de entrada de la red autoasociativa equivalente. También puede conseguirse que una red autoasociativa se comporte como heteroasociativa, simplemente presentando, como entrada parcial de la autoasociativa, la información de entrada para la

heteroasociativa y haciendo que la red complete la información para producir lo que sería la salida de la red heteroasociativa equivalente.

2.5.2.5 Redes heteroasociativas

Las redes heteroasociativas, al asociar informaciones de entrada con diferentes informaciones de salida, precisan al menos de dos capas, una para captar y retener la información de entrada y otra para mantener la salida con la información asociada. Si esto no fuese así, se perdería la información inicial al obtenerse el dato asociado {3}, lo cual no debe ocurrir, ya que en el proceso de obtención de la salida se puede necesitar acceder varias veces a esta información que, por tanto, deberá permanecer en la capa de entrada.

En cuanto a su conectividad, pueden ser del tipo con conexión hacia adelante (o feedforward) o con conexión hacia atrás (feedforward/feedback), o bien con conexiones laterales.

2.5.2.6 Redes autoasociativas

Una red autoasociativa asocia una información de entrada con el ejemplar más parecido de los almacenados conocidos por la red. Estos tipos de redes pueden implementarse con una sola capa de neuronas. Esta capa comenzará reteniendo la información inicial a la entrada, y terminará representando la información autoasociada. Si se quiere mantener la información de entrada y salida, se deberían añadir capas adicionales, sin embargo, la funcionalidad de la red puede conseguirse en una sola capa.

En cuanto a su conectividad, existen de conexiones laterales y, en algunos casos, conexiones autorrecurrentes.

2.5.3 Mecanismos de aprendizaje para redes neuronales

Se ha visto que los datos de entrada se procesan a través de la red neuronal con el propósito de lograr una salida. También se dijo que las redes neuronales extraen generalizaciones desde un conjunto determinado de ejemplos anteriores de tales problemas de decisión. Una red neuronal debe aprender a calcular la salida correcta para cada constelación (arreglo o vector) de entrada en el conjunto de ejemplos. Este proceso de aprendizaje se denomina: proceso de entrenamiento o acondicionamiento. El conjunto de datos (o conjunto de ejemplos) sobre el cual este proceso se basa es llamado: conjunto de datos de entrenamiento.

Si la topología de la red y las diferentes funciones de cada neurona (entrada, activación y salida) no pueden cambiar durante el aprendizaje, mientras que los pesos sobre cada una de las conexiones si pueden hacerlo; el aprendizaje de una red neuronal significa: adaptación de los pesos.

En otras palabras el aprendizaje es el proceso por el cual una red neuronal modifica sus pesos en respuesta a una información de entrada. Los cambios que se producen durante el mismo se reducen a la destrucción, modificación y creación de conexiones entre las neuronas. En los sistemas biológicos existe una continua destrucción y creación de conexiones entre las neuronas. En los modelos de redes neuronales artificiales, la creación de una nueva conexión implica que el peso de la misma pasa a tener un valor distinto de cero. De la misma manera, una conexión se destruye cuando su peso pasa a ser cero.

Durante el proceso de aprendizaje, los pesos de las conexiones de la red sufren modificaciones, por lo tanto, se puede afirmar que este proceso ha terminado (la red ha aprendido) cuando los valores de los pesos permanecen estables ($dw_{ij}/dt = 0$).

Un aspecto importante respecto al aprendizaje de las redes neuronales es el conocer cómo se modifican los valores de los pesos, es decir, cuáles son los criterios que se siguen para cambiar el valor asignado a las conexiones cuando se pretende que la red aprenda una nueva información.

Hay dos métodos de aprendizaje importantes que pueden distinguirse: aprendizaje supervisado y aprendizaje no supervisado.

2.5.3.1 Aprendizaje supervisado

El aprendizaje supervisado se caracteriza porque el proceso de aprendizaje se realiza mediante un entrenamiento controlado por un agente externo (supervisor, maestro) que determina la respuesta que debería generar la red a partir de una entrada determinada. El supervisor controla la salida de la red y en caso de que ésta no coincida con la deseada, se procederá a modificar los pesos de las conexiones, con el fin de conseguir que la salida obtenida se aproxime a la deseada. En este tipo de aprendizaje se suelen considerar, a su vez, tres formas de llevarlo a cabo, que dan lugar a los siguientes aprendizajes supervisados:

- 1) Aprendizaje por corrección de error: Consiste en ajustar los pesos de las conexiones de la red en función de la diferencia entre los valores deseados y los obtenidos a la salida de la red, es decir, en función del error cometido en la salida. Un ejemplo de este tipo de algoritmos lo constituye la regla de aprendizaje del Perceptron, utilizada en el entrenamiento de la red del mismo nombre que desarrolló Rosenblatt en 1958 (Rosenblatt, 1958). Esta es una regla muy simple, para cada neurona en la capa de salida se le calcula la desviación a la salida objetivo como el error, δ . El cual luego se utiliza para cambiar los pesos sobre la conexión de la neurona precedente. El cambio de los

pesos por medio de la regla de aprendizaje del Perceptron se realiza según la siguiente ecuación:

$$\Delta W_{ij} = \sigma * out_j * (a_{qi} - out_i) \quad (38)$$

Donde a_{qi} es la salida deseada de la neurona de salida N_i , $\delta_i = (a_{qi} - out_i)$ la desviación objetivo de la neurona N_i y σ es el coeficiente de aprendizaje. Otro algoritmo muy conocido y que pertenece a esta clasificación es la regla de aprendizaje Delta o regla del mínimo error cuadrado (LMS Error: Least Mean Squared Error), que también utiliza la desviación a la salida objetivo, pero toma en consideración a todas las neuronas predecesoras que tiene la neurona de salida. Esto permite cuantificar el error global cometido en cualquier momento durante el proceso de entrenamiento de la red, lo cual es importante, ya que cuanto más información se tenga sobre el error cometido, más rápido se puede aprender. Luego el error calculado (δ) es igualmente repartido entre las conexiones de las neuronas predecesoras.

- 2) Aprendizaje por refuerzo: Se trata de un aprendizaje supervisado, más lento que el anterior, que se basa en la idea de no disponer de un ejemplo completo del comportamiento deseado, es decir, de no indicar durante el entrenamiento exactamente la salida que se desea que proporcione la red ante una determinada entrada. En el aprendizaje por refuerzo la función del supervisor se reduce a indicar mediante una señal de refuerzo si la salida obtenida en la red se ajusta a la deseada (éxito = +1 o fracaso = -1), y en función de ello se ajustan los pesos basándose en un mecanismo de probabilidades. Se podría decir que en este tipo de aprendizaje la función del supervisor se asemeja más a la de un crítico (que opina sobre la respuesta de la red) que a la de un maestro (que indica a la

red la respuesta concreta que debe generar), como ocurría en el caso de supervisión por corrección del error.

- 3) Aprendizaje estocástico: Consiste básicamente en realizar cambios aleatorios en los valores de los pesos de las conexiones de la red y evaluar su efecto a partir del objetivo deseado y de distribuciones de probabilidad. En el aprendizaje estocástico se suele hacer una analogía en términos termodinámicos, asociando a la red neuronal con un sólido físico que tiene cierto estado energético. En el caso de la red, la energía de la misma representaría el grado de estabilidad de la red, de tal forma que el estado de mínima energía correspondería a una situación en la que los pesos de las conexiones consiguen que su funcionamiento sea el que más se ajusta al objetivo deseado. Según lo anterior, el aprendizaje consistiría en realizar un cambio aleatorio de los valores de los pesos y determinar la energía de la red (habitualmente la función energía es una función de Liapunov). Si la energía es menor después del cambio, es decir, si el comportamiento de la red se acerca al deseado, se acepta el cambio; si, por el contrario, la energía no es menor, se aceptaría el cambio en función de una determinada y preestablecida distribución de probabilidades.

2.5.3.2 Aprendizaje no supervisado

Las redes con aprendizaje no supervisado (también conocido como autosupervisado) no requieren influencia externa para ajustar los pesos de las conexiones entre sus neuronas. La red no recibe ninguna información por parte del entorno que le indique si la salida generada en respuesta a una determinada entrada es o no correcta.

Estas redes deben encontrar las características, regularidades, correlaciones o categorías que se puedan establecer entre los datos que se presenten en su entrada. Existen varias posibilidades en cuanto a la interpretación de la salida de estas redes, que dependen de su estructura y del algoritmo de aprendizaje empleado.

En algunos casos, la salida representa el grado de familiaridad o similitud entre la información que se le está presentando en la entrada y las informaciones que se le han mostrado hasta entonces (en el pasado). En otro caso, podría realizar una clusterización (clustering) o establecimiento de categorías, indicando la red a la salida a qué categoría pertenece la información presentada a la entrada, siendo la propia red quien debe encontrar las categorías apropiadas a partir de las correlaciones entre las informaciones presentadas. En cuanto a los algoritmos de aprendizaje no supervisado, en general se suelen considerar dos tipos, que dan lugar a los siguientes aprendizajes:

- 1) Aprendizaje hebbiano: Esta regla de aprendizaje es la base de muchas otras, la cual pretende medir la familiaridad o extraer características de los datos de entrada. El fundamento es una suposición bastante simple: si dos neuronas N_i y N_j toman el mismo estado simultáneamente (ambas activas o ambas inactivas), el peso de la conexión entre ambas se incrementa. Las entradas y salidas permitidas a la neurona son: $\{-1, 1\}$ o $\{0, 1\}$ (neuronas binarias). Esto puede explicarse porque la regla de aprendizaje de Hebb se originó a partir de la neurona biológica clásica, que solamente puede tener dos estados: activa o inactiva.
- 2) Aprendizaje competitivo y comparativo: Se orienta a la clusterización o clasificación de los datos de entrada. Como característica principal del aprendizaje competitivo se puede decir que, si un patrón nuevo

se determina que pertenece a una clase reconocida previamente, entonces la inclusión de este nuevo patrón a esta clase matizará la representación de la misma. Si el patrón de entrada se determinó que no pertenece a ninguna de las clases reconocidas anteriormente, entonces la estructura y los pesos de la red neuronal serán ajustados para reconocer la nueva clase.

3. Metodología

A lo largo de este capítulo se describirá con detalle los pasos que fueron necesarios para desarrollar este proyecto. El punto de inicio es conocer el comportamiento de la población de estudio y esto se presenta a continuación.

3.1 Muestreo de reconocimiento

El lugar donde se realizó el muestreo de reconocimiento fue el establo del campus Amazcala de la Universidad Autónoma de Querétaro, en la figura 3.1 se muestra la sala de ordeño la cual es de tipo espina de pescado y tiene capacidad para ordeñar cuatro vacas simultáneamente.

La toma de muestras de la leche se realizó por medio del uso de un pesador de leche, el cual fue instalado en las tuberías por donde viaja la leche que está siendo recolectada, en la figura 3.2 se muestra un pesador de leche. Este equipo es de gran utilidad ya que recolecta un pequeño porcentaje de la cantidad total de leche así como también nos permite obtener las muestras durante el proceso de ordeña con esto es posible obtener muestras de todo el espectro de leche disponible.



FIGURA 3.1, Sala de ordeño del campus Amazcala.



FIGURA 3.2, Pesador de leche.

Cada muestra debe de ser de al menos 70 ml para poder ser analizada posteriormente por medio de un LactoscanSP, figura 3.3, el cual tiene una precisión de; $\pm 0.1\%$ de contenido de grasa, $\pm 0.2\%$ de contenido de lactosa, $\pm 0.15\%$ de contenido de proteínas, $\pm 0.3 \text{ Kg/m}^3$ de densidad, $\pm 0.05 \text{ mS}$ de conductividad y $\pm 1 \text{ }^\circ\text{C}$ de temperatura. Con esto tendremos las variables de entrada y salida para entrenar la red neuronal. En la figura 3.4 se presentan las muestras que fueron analizadas en el laboratorio de lácteos del campus Amazcala.



FIGURA 3.3, LactoscanSP.



FIGURA 3.4, Muestras correspondientes al muestro de reconocimiento.

Los resultados del análisis se presentan en la tabla 3.1 en la cual se presentan las variables contenido de grasa, contenido de lactosa, contenido de proteína, densidad, conductividad y temperatura de 33 muestras. Por medio de los valores obtenidos en esta tabla se podrán realizar estimaciones con respecto a la población de estudio.

TABLA 3.1, Resultados del análisis para la muestra de reconocimiento.

Numero de muestra	Contenido grasa (%)	Contenido lactosa (%)	Contenido proteína (%)	Densidad (Kg/m3)	Conductividad Eléctrica (mS)	Temperatura (°C)
1	3.97	4.64	3.26	1030.28	3.97	28.2
2	2.57	4.16	2.92	1027.51	4	28.5
3	3.52	4.03	2.83	1026.29	4.01	29
4	3.49	4.07	2.86	1026.55	4.01	29.5
5	3.45	4.11	2.89	1026.85	4	29.3
6	3.36	4.1	2.88	1026.8	4.01	29.5
7	3.57	4.05	2.85	1026.43	4.01	29.6
8	3.82	4.17	2.93	1027.13	4.01	30.2
9	2.21	4.47	3.13	1029.78	4.08	29.9
10	2.4	4.41	3.09	1029.31	4.11	29.4
11	4.03	4.22	2.97	1027.39	4.16	30.1
12	3.44	4.84	3.4	1031.86	4.04	30
13	2.21	4.53	3.17	1030.16	4	29.3
14	2.57	4.49	3.15	1029.77	4.16	29.5
15	3.14	4.53	3.18	1029.86	4.15	29.6
16	1.73	4.72	3.31	1031.67	3.98	29.8
17	3.4	4.77	3.35	1031.39	4.1	29.5
18	4.07	4.73	3.32	1030.86	4.11	29.6
19	4.93	4.74	3.33	1030.6	4.12	30
20	4.44	4.5	3.17	1029.15	4.07	30.6
21	3.04	4.32	3.03	1028.45	4.17	29.9
22	3.71	4.37	3.07	1028.58	4.15	30.1
23	4.93	4.37	3.07	1028.06	4.17	30.3
24	2.55	4.36	3.06	1028.9	3.96	30.3
25	3.18	4.4	3.09	1028.98	4.02	30.4
26	3.3	4.29	3.02	1028.17	4.03	30.2
27	3.33	4.25	2.99	1027.88	4.03	30.3
28	4.4	4.4	3.1	1028.49	4.03	30.6
29	2.65	4.63	3.25	1030.7	4.01	31.2
30	2.37	4.58	3.21	1030.51	4	31
31	2.27	4.47	3.13	1029.76	4	30.8
32	2.3	4.5	3.15	1029.94	3.99	30.6
33	2.43	4.44	3.11	1029.49	4.01	30.7

3.2 Prueba de normalidad

Una vez que contamos con los datos de la prueba de reconocimiento, tabla 3.1, realizaremos una prueba de normalidad para los datos de contenido de grasa, contenido de lactosa y contenido de proteínas. Esto lo realizaremos por medio de una prueba de Anderson-Darling la cual se ve representada por la ecuación 26, se selecciono esta prueba ya que es fácil de implementar y es ampliamente utilizada por la comunidad científica.

En la tabla 3.2 se presentan los resultados de la prueba de normalidad para la variable contenido de grasa, en la tabla 3.3 se presentan los resultados de la prueba de normalidad para la variable contenido de lactosa y en la tabla 3.4 se presentan los resultados de la prueba de normalidad para la variable contenido de proteína.

Los resultados de la prueba de normalidad son para el contenido de grasa $A^2=0.447$, para el contenido de lactosa $A^2=0.25$ y para el contenido de proteína $A^2=0.239$. La interpretación de estos resultados es la siguiente, podemos afirmar con base a estas pruebas de normalidad que las tres variables analizadas tienden a una distribución normal con un 95% de probabilidad. De los resultados de esta prueba también podemos afirmar que la variable contenido de grasa será la variable dominante del modelo ya que esta presento el coeficiente A^2 mas alto de las tres variables analizadas. Esto quiere decir que la variable grasa será la que presentara más dificultad para ser modelada seguida por la lactosa mientras que la variable proteína será la que se podrá modelar con más facilidad.

Cabe señalar que si comparamos los datos de la tabla 1.1 con los datos obtenidos durante esta prueba se puede observar que tienen propiedades similares tal es el caso de la media aritmética la cual se asemeja bastante con la de los datos obtenidos. Es por esto que se puede afirmar que los datos obtenidos en esta prueba son representativos de la población de estudio ya que los datos de la tabla 1.1 fueron obtenidos a partir del estudio de Pinto et al. (1998).

TABLA 3.2, Prueba de normalidad para la variable contenido de grasa.

i	X(i)	X(i) Ord	Z(i)	F(Z(i))	1-F(Z(i))	Ord col 6	Si
1	1.730	1.730	-1.841	0.033	0.967	0.019	-7.372
2	2.210	2.210	-1.254	0.105	0.895	0.019	-18.629
3	2.210	2.210	-1.254	0.105	0.895	0.070	-24.537
4	2.270	2.270	-1.181	0.119	0.881	0.077	-32.830
5	2.300	2.300	-1.144	0.126	0.874	0.154	-35.467
6	2.370	2.370	-1.059	0.145	0.855	0.166	-41.018
7	2.400	2.400	-1.022	0.153	0.847	0.185	-46.328
8	2.430	2.430	-0.985	0.162	0.838	0.238	-48.840
9	2.550	2.550	-0.838	0.201	0.799	0.281	-48.864
10	2.570	2.570	-0.814	0.208	0.792	0.341	-50.270
11	2.570	2.570	-0.814	0.208	0.792	0.364	-54.209
12	2.650	2.650	-0.716	0.237	0.763	0.378	-55.497
13	3.040	3.040	-0.239	0.405	0.595	0.397	-45.686
14	3.140	3.140	-0.117	0.453	0.547	0.401	-46.002
15	3.180	3.180	-0.068	0.473	0.527	0.420	-46.850
16	3.300	3.300	0.079	0.531	0.469	0.440	-45.082
17	3.330	3.330	0.115	0.546	0.454	0.454	-46.027
18	3.360	3.360	0.152	0.560	0.440	0.469	-46.794
19	3.400	3.400	0.201	0.580	0.420	0.527	-43.870
20	3.440	3.440	0.250	0.599	0.401	0.547	-43.571
21	3.450	3.450	0.262	0.603	0.397	0.595	-42.034
22	3.490	3.490	0.311	0.622	0.378	0.763	-32.042
23	3.520	3.520	0.348	0.636	0.364	0.792	-30.856
24	3.570	3.570	0.409	0.659	0.341	0.792	-30.578
25	3.710	3.710	0.580	0.719	0.281	0.799	-27.154
26	3.820	3.820	0.714	0.762	0.238	0.838	-22.860
27	3.970	3.970	0.898	0.815	0.185	0.847	-19.648
28	4.030	4.030	0.971	0.834	0.166	0.855	-18.578
29	4.070	4.070	1.020	0.846	0.154	0.874	-17.219
30	4.400	4.400	1.423	0.923	0.077	0.881	-12.212
31	4.440	4.440	1.472	0.930	0.070	0.895	-11.217
32	4.930	4.930	2.071	0.981	0.019	0.895	-8.200
33	4.930	4.930	2.071	0.981	0.019	0.967	-3.426
media	3.236					suma	-1103.766
des est	0.818					A²	0.447

TABLA 3.3, Prueba de normalidad para la variable contenido de lactosa.

i	X(i)	X(i) Ord	Z(i)	F(Z(i))	1-F(Z(i))	Ord col 6	Si
1	4.640	4.030	-1.719	0.043	0.957	0.028	-6.717
2	4.160	4.050	-1.629	0.052	0.948	0.055	-17.565
3	4.030	4.070	-1.540	0.062	0.938	0.072	-27.058
4	4.070	4.100	-1.405	0.080	0.920	0.079	-35.490
5	4.110	4.110	-1.361	0.087	0.913	0.085	-44.144
6	4.100	4.160	-1.137	0.128	0.872	0.156	-43.080
7	4.050	4.170	-1.092	0.137	0.863	0.167	-49.089
8	4.170	4.220	-0.868	0.193	0.807	0.229	-46.837
9	4.470	4.250	-0.734	0.232	0.768	0.302	-45.244
10	4.410	4.290	-0.555	0.290	0.710	0.302	-46.320
11	4.220	4.320	-0.421	0.337	0.663	0.350	-44.882
12	4.840	4.360	-0.241	0.405	0.595	0.350	-44.955
13	4.530	4.370	-0.197	0.422	0.578	0.367	-46.644
14	4.490	4.370	-0.197	0.422	0.578	0.401	-47.969
15	4.530	4.400	-0.062	0.475	0.525	0.401	-48.087
16	4.720	4.400	-0.062	0.475	0.525	0.454	-47.579
17	4.770	4.410	-0.018	0.493	0.507	0.507	-45.754
18	4.730	4.440	0.117	0.546	0.454	0.525	-43.713
19	4.740	4.470	0.251	0.599	0.401	0.525	-42.808
20	4.500	4.470	0.251	0.599	0.401	0.578	-41.364
21	4.320	4.490	0.340	0.633	0.367	0.578	-41.210
22	4.370	4.500	0.385	0.650	0.350	0.595	-40.822
23	4.370	4.500	0.385	0.650	0.350	0.663	-37.885
24	4.360	4.530	0.520	0.698	0.302	0.710	-32.943
25	4.400	4.530	0.520	0.698	0.302	0.768	-30.500
26	4.290	4.580	0.743	0.771	0.229	0.807	-24.154
27	4.250	4.630	0.967	0.833	0.167	0.863	-17.502
28	4.400	4.640	1.012	0.844	0.156	0.872	-16.836
29	4.630	4.720	1.370	0.915	0.085	0.913	-10.262
30	4.580	4.730	1.415	0.921	0.079	0.920	-9.745
31	4.470	4.740	1.460	0.928	0.072	0.938	-8.465
32	4.500	4.770	1.594	0.945	0.055	0.948	-6.937
33	4.440	4.840	1.907	0.972	0.028	0.957	-4.708
media	4.414					suma	-1097.268
des est	0.223					A²	0.251

TABLA 3.4, Prueba de normalidad para la variable contenido de proteína.

i	X(i)	X(i) Ord	Z(i)	F(Z(i))	1-F(Z(i))	Ord col 6	Si
1	3.260	2.830	-1.733	0.042	0.958	0.026	-6.817
2	2.920	2.850	-1.604	0.054	0.946	0.053	-17.542
3	2.830	2.860	-1.539	0.062	0.938	0.069	-27.317
4	2.860	2.880	-1.411	0.079	0.921	0.077	-35.659
5	2.890	2.890	-1.346	0.089	0.911	0.087	-43.714
6	2.880	2.920	-1.153	0.124	0.876	0.150	-43.786
7	2.850	2.930	-1.089	0.138	0.862	0.166	-49.109
8	2.930	2.970	-0.831	0.203	0.797	0.238	-45.482
9	3.130	2.990	-0.702	0.241	0.759	0.301	-44.575
10	3.090	3.020	-0.509	0.305	0.695	0.324	-43.957
11	2.970	3.030	-0.445	0.328	0.672	0.324	-47.064
12	3.400	3.060	-0.252	0.401	0.599	0.372	-43.811
13	3.170	3.070	-0.187	0.426	0.574	0.372	-46.103
14	3.150	3.070	-0.187	0.426	0.574	0.421	-46.408
15	3.180	3.090	-0.059	0.477	0.523	0.421	-46.567
16	3.310	3.090	-0.059	0.477	0.523	0.472	-46.243
17	3.350	3.100	0.006	0.502	0.498	0.498	-45.748
18	3.320	3.110	0.070	0.528	0.472	0.523	-45.017
19	3.330	3.130	0.199	0.579	0.421	0.523	-44.185
20	3.170	3.130	0.199	0.579	0.421	0.574	-42.950
21	3.030	3.150	0.328	0.628	0.372	0.574	-41.783
22	3.070	3.150	0.328	0.628	0.372	0.599	-41.984
23	3.070	3.170	0.457	0.676	0.324	0.672	-35.521
24	3.060	3.170	0.457	0.676	0.324	0.695	-35.523
25	3.090	3.180	0.521	0.699	0.301	0.759	-31.087
26	3.020	3.210	0.714	0.762	0.238	0.797	-25.402
27	2.990	3.250	0.972	0.834	0.166	0.862	-17.475
28	3.100	3.260	1.036	0.850	0.150	0.876	-16.253
29	3.250	3.310	1.358	0.913	0.087	0.911	-10.522
30	3.210	3.320	1.422	0.923	0.077	0.921	-9.623
31	3.130	3.330	1.487	0.931	0.069	0.938	-8.225
32	3.150	3.350	1.616	0.947	0.053	0.946	-6.959
33	3.110	3.400	1.937	0.974	0.026	0.958	-4.496
media	3.099					suma	-1096.909
des est	0.155					A²	0.240

3.3 Determinación del tamaño de la muestra según los métodos tradicionales

Para la determinación del tamaño de la muestra haremos uso de los datos de la variable grasa ya que como se menciona anteriormente es la variable dominante en el modelo.

La varianza muestral de la variable contenido de grasa tiene un valor de $S^2=0.6689$ y el error permisible para el modelo será $\varepsilon=0.0815$ lo que representa un 5% de la escala total de medición para la variable grasa. Se trabajara con un nivel de confianza del 95% por lo que $\alpha=0.05$ y por lo tanto las abscisas que cubren este intervalo de probabilidad tienen un valor de $z_{0.025}=1.96$. Finalmente se utilizara la ecuación 29 y se sustituyen los datos anteriores.

$$n = z_{\alpha/2}^2 \frac{\sigma^2}{\varepsilon^2} = 1.96^2 \frac{0.6689}{0.0815^2} = 386.86 \approx 387 \quad (39)$$

De la ecuación 39 se obtiene que el numero de muestras necesarias para cumplir con los limites que fijamos es de 386.86 muestras lo cual redondeamos a 387 muestras. Nótese que si se hubiera realizado esta determinación del tamaño de la muestra con la variable contenido de proteína, el numero de muestras necesarias seria solo de 14 muestras ya que su varianza muestral es $S^2=0.0241$.

3.4 Desarrollo de los mecanismos de aprendizaje

En esta sección se desarrollaran los algoritmos de aprendizaje ya que estos son diferentes dependiendo de la topología y la función de activación que este siendo utilizada en cada red neuronal. La topología de las redes será de tres entradas, en la capa oculta el numero de neuronas variara desde una hasta quince y en la capa de salida se tendrán 3 neuronas.

3.4.1 Mecanismo de aprendizaje función de activación Lineal

Se propone la función de error de la ecuación 40, el error durante el entrenamiento a la salida de las neuronas de la capa de salida se representa por la ecuación 41, la ecuación 42 representa la salida de la función de activación de las neuronas de la capa de salida, la ecuación 43 representa la función de entrada a las neuronas de la capa de salida, la ecuación 44 representa la salida de la función de activación de las neuronas en la capa oculta y la ecuación 45 representa la función de entrada de las neuronas en la capa oculta.

$$E = \frac{1}{2} \sum_{k=1}^K e_k^2 \quad (40)$$

$$e_k = Y_k - O_k \quad (41)$$

$$O_k = A \cdot S_k + B \quad (42)$$

$$S_k = \sum_{j=1}^J O_j \cdot W_{jk} \quad (43)$$

$$O_j = A \cdot S_j + B \quad (44)$$

$$S_j = \sum_{i=1}^I X_i \cdot W_{ij} \quad (45)$$

El propósito de un mecanismo de aprendizaje es obtener un nuevo valor de peso de interconexión que mejore la salida de la red, por que tanto se desea minimizar la función de error, ecuación 40, con respecto a los pesos de interconexión y este se representa en la ecuación 46, por medio de la regla de la cadena podemos ir avanzando sobre las ecuaciones 40-45. Por lo tanto las ecuaciones de ajuste para los pesos de interconexión se muestran en las ecuaciones 47 y 48.

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial W_{ij}} \\ \frac{\partial E}{\partial W_{jk}} \end{bmatrix} = \begin{bmatrix} \frac{\partial E}{\partial e_k} \frac{\partial e_k}{\partial O_k} \frac{\partial O_k}{\partial S_k} \frac{\partial S_k}{\partial O_j} \frac{\partial O_j}{\partial S_j} \frac{\partial S_j}{\partial W_{ij}} \\ \frac{\partial E}{\partial e_k} \frac{\partial e_k}{\partial O_k} \frac{\partial O_k}{\partial S_k} \frac{\partial S_k}{\partial W_{jk}} \end{bmatrix} \Rightarrow \quad (46)$$

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial W_{ij}} \\ \frac{\partial E}{\partial W_{jk}} \end{bmatrix} = \begin{bmatrix} (e_k)(-1)(A)(W_{jk})(A)(X_i) \\ (e_k)(-1)(A)(O_j) \end{bmatrix}$$

$$W_{ij}(t) = W_{ij}(t-1) + A^2 e_k X_i W_{jk} \quad (47)$$

$$W_{jk}(t) = W_{jk}(t-1) + A e_k O_j \quad (48)$$

3.4.2 Mecanismo de aprendizaje función de activación Wavelet

Ya que la obtención de las ecuaciones de aprendizaje es similar en todos los caso, solo se hará hincapié en las diferencias que se tienen. La ecuación 49 representa la salida de la función de activación de las neuronas de la capa de salida, la ecuación 50 representa la función de entrada a las neuronas de la capa de salida, la ecuación 51 representa la salida de la función de activación de las neuronas en la capa oculta y la ecuación 52 representa la función de entrada de las neuronas en la capa oculta.

$$O_k = (1 - S_k^2) \cdot e^{(-S_k^2/2)} \quad (49)$$

$$S_k = \sum_{j=1}^J O_j \cdot W_{jk} \quad (50)$$

$$O_j = (1 - S_j^2) \cdot e^{(-S_j^2/2)} \quad (51)$$

$$S_j = \sum_{i=1}^I X_i \cdot W_{ij} \quad (52)$$

El propósito de un mecanismo de aprendizaje es obtener un nuevo valor de peso de interconexión que mejore la salida de la red, por que tanto se desea minimizar la función de error, ecuación 40, con respecto a los pesos de interconexión y este se representa en la ecuación 53, por medio de la regla de la cadena podemos ir avanzando sobre las ecuaciones 40-41 y 49-52. Por lo tanto las ecuaciones de ajuste para los pesos de interconexión se muestran en las ecuaciones 54 y 55.

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial W_{ij}} \\ \frac{\partial E}{\partial W_{jk}} \end{bmatrix} = \begin{bmatrix} (e_k)(-1)(S_k \cdot e^{-S_k^2/2} \cdot (S_k - 3))(W_{jk})(S_j \cdot e^{-S_j^2/2} \cdot (S_j - 3))(X_i) \\ (e_k)(-1)(S_k \cdot e^{-S_k^2/2} \cdot (S_k - 3))(O_j) \end{bmatrix} \quad (53)$$

$$W_{ij}(t) = W_{ij}(t-1) + (e_k)(S_k \cdot e^{-S_k^2/2} \cdot (S_k - 3))(W_{jk})(S_j \cdot e^{-S_j^2/2} \cdot (S_j - 3))(X_i) \quad (54)$$

$$W_{jk}(t) = W_{jk}(t-1) + (e_k)(S_k \cdot e^{-S_k^2/2} \cdot (S_k - 3))(O_j) \quad (55)$$

3.4.3 Mecanismo de aprendizaje función de activación Sigmoidea

Ya que la obtención de las ecuaciones de aprendizaje es similar en todos los caso, solo se hará hincapié en las diferencias que se tienen. La ecuación 56 representa la salida de la función de activación de las neuronas de la capa de salida, la ecuación 57 representa la función de entrada a las neuronas de la capa de salida, la ecuación 58 representa la salida de la función de activación de las neuronas en la capa oculta y la ecuación 59 representa la función de entrada de las neuronas en la capa oculta.

$$O_k = \frac{1}{1 + e^{-S_k}} \quad (56)$$

$$S_k = \sum_{j=1}^J O_j \cdot W_{jk} \quad (57)$$

$$O_j = \frac{1}{1 + e^{-S_j}} \quad (58)$$

$$S_j = \sum_{i=1}^I X_i \cdot W_{ij} \quad (59)$$

El propósito de un mecanismo de aprendizaje es obtener un nuevo valor de peso de interconexión que mejore la salida de la red, por que tanto se desea minimizar la función de error, ecuación 40, con respecto a los pesos de interconexión y este se representa en la ecuación 60, por medio de la regla de la cadena podemos ir avanzando sobre las ecuaciones 40-41 y 56-59. Por lo tanto las ecuaciones de ajuste para los pesos de interconexión se muestran en las ecuaciones 61 y 62.

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial W_{ij}} \\ \frac{\partial E}{\partial W_{jk}} \end{bmatrix} = \begin{bmatrix} (e_k)(-1)(e^{-S_k}[e^{-S_k} + 1]^{-2})(W_{jk})(e^{-S_j}[e^{-S_j} + 1]^{-2})(X_i) \\ (e_k)(-1)(e^{-S_k}[e^{-S_k} + 1]^{-2})(O_j) \end{bmatrix} \quad (60)$$

$$W_{ij}(t) = W_{ij}(t-1) + (e_k)(e^{-S_k}[e^{-S_k} + 1]^{-2})(W_{jk})(e^{-S_j}[e^{-S_j} + 1]^{-2})(X_i) \quad (61)$$

$$W_{jk}(t) = W_{jk}(t-1) + (e_k)(e^{-S_k}[e^{-S_k} + 1]^{-2})(O_j) \quad (62)$$

3.4.4 Mecanismo de aprendizaje función de activación Tangente hiperbólica

Ya que la obtención de las ecuaciones de aprendizaje es similar en todos los caso, solo se hará hincapié en las diferencias que se tienen. La ecuación 63 representa la salida de la función de activación de las neuronas de la capa de salida, la ecuación 64 representa la función de entrada a las neuronas de la capa de salida, la ecuación 65 representa la salida de la función de activación de las neuronas en la capa oculta y la ecuación 66 representa la función de entrada de las neuronas en la capa oculta.

$$O_k = \frac{2}{1 + e^{-2S_k}} - 1 \quad (63)$$

$$S_k = \sum_{j=1}^J O_j \cdot W_{jk} \quad (64)$$

$$O_j = \frac{2}{1 + e^{-2S_j}} - 1 \quad (65)$$

$$S_j = \sum_{i=1}^I X_i \cdot W_{ij} \quad (66)$$

El propósito de un mecanismo de aprendizaje es obtener un nuevo valor de peso de interconexión que mejore la salida de la red, por que tanto se desea minimizar la función de error, ecuación 40, con respecto a los pesos de interconexión y este se representa en la ecuación 67, por medio de la regla de la cadena podemos ir avanzando sobre las ecuaciones 40-41 y 63-66. Por lo tanto las ecuaciones de ajuste para los pesos de interconexión se muestran en las ecuaciones 68 y 69.

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial W_{ij}} \\ \frac{\partial E}{\partial W_{jk}} \end{bmatrix} = \begin{bmatrix} (e_k)(-1)(4e^{-2S_k}[1+e^{-2S_k}]^{-2})(W_{jk})(4e^{-2S_j}[1+e^{-2S_j}]^{-2})(X_i) \\ (e_k)(-1)(4e^{-2S_k}[1+e^{-2S_k}]^{-2})(O_j) \end{bmatrix} \quad (67)$$

$$W_{ij}(t) = W_{ij}(t-1) + (e_k)(4e^{-2S_k}[1+e^{-2S_k}]^{-2})(W_{jk})(4e^{-2S_j}[1+e^{-2S_j}]^{-2})(X_i) \quad (68)$$

$$W_{jk}(t) = W_{jk}(t-1) + (e_k)(4e^{-2S_k}[1+e^{-2S_k}]^{-2})(O_j) \quad (69)$$

3.4.5 Mecanismo de aprendizaje función de activación Gaussiana

Ya que la obtención de las ecuaciones de aprendizaje es similar en todos los casos, solo se hará hincapié en las diferencias que se tienen. La ecuación 70 representa la salida de la función de activación de las neuronas de la capa de salida, la ecuación 71 representa la función de entrada a las neuronas de la capa de salida, la ecuación 72 representa la salida de la función de activación de las neuronas en la capa oculta y la ecuación 73 representa la función de entrada de las neuronas en la capa oculta.

$$O_k = e^{-S_k^2} \quad (70)$$

$$S_k = \sum_{j=1}^J O_j \cdot W_{jk} \quad (71)$$

$$O_j = e^{-S_j^2} \quad (72)$$

$$S_j = \sum_{i=1}^I X_i \cdot W_{ij} \quad (73)$$

El propósito de un mecanismo de aprendizaje es obtener un nuevo valor de peso de interconexión que mejore la salida de la red, por que tanto se desea minimizar la función de error, ecuación 40, con respecto a los pesos de interconexión y este se representa en la ecuación 74, por medio de la regla de la cadena podemos ir avanzando sobre las ecuaciones 40-41 y 70-73. Por lo tanto las ecuaciones de ajuste para los pesos de interconexión se muestran en las ecuaciones 75 y 76.

$$\nabla E = \begin{bmatrix} \frac{\partial E}{\partial W_{ij}} \\ \frac{\partial E}{\partial W_{jk}} \end{bmatrix} = \begin{bmatrix} (e_k)(-1)(-2S_k e^{-S_k^2})(W_{jk})(-2S_j e^{-S_j^2})(X_i) \\ (e_k)(-1)(-2S_k e^{-S_k^2})(O_j) \end{bmatrix} \quad (67)$$

$$W_{ij}(t) = W_{ij}(t-1) + (e_k)(-2S_k e^{-S_k^2})(W_{jk})(-2S_j e^{-S_j^2})(X_i) \quad (68)$$

$$W_{jk}(t) = W_{jk}(t-1) + (e_k)(-2S_k e^{-S_k^2})(O_j) \quad (69)$$

3.5 Método de selección de la mejor topología de red neuronal

El procedimiento que se realizo para seleccionar la topología de la red neuronal y su función de activación está basado en la aplicación de la ecuación 70 la cual es propuesta en este trabajo como un método con el que podamos justificar la elección de la red neuronal más apropiada. En este caso las redes neuronales fueron entrenadas durante 100,000 épocas con los datos obtenidos en el muestreo de reconocimiento, tabla 3.1. Los resultados se presentan en las tablas 3.5 a 3.9.

$$T = |e_{\max} \cdot B \cdot E / r| \quad (70)$$

TABLA 3.5, Selección de topología para la función de activación Lineal.

Numero de neuronas en la capa oculta	e_{max}	r	B	E	$T= e_{max} * B * E / r $
1	0.0628	0.3088	-0.0129	-7.604	0.01994868
2	0.0624	0.318	-0.0124	-8.7692	0.02133728
3	0.0621	0.3139	-0.0119	-9.604	0.02260994
4	0.0637	0.2451	-0.011	-11.9162	0.03406643
5	0.0651	0.2605	-0.0115	-9.8909	0.02842542
6	0.0633	0.31	-0.0109	-10.528	0.02343227
7	0.066	0.1721	-0.0119	-12.1583	0.05548593
8	0.0658	0.2385	-0.0105	-11.8783	0.03440972
9	0.0617	0.3031	-0.0126	-12.0587	0.03092931
10	0.0678	0.1054	-0.0103	-14.3448	0.09504315
11	0.0678	0.1913	-0.0096	-13.2025	0.04492025
12	0.0663	0.0314	-0.0106	-17.4861	0.39136565
13	0.066	-0.0394	-0.0089	-18.2602	0.27223456
14	0.0574	0.2425	-0.0081	-13.6205	0.02611429
15	0.0585	0.205	-0.0058	-11.9286	0.01974329

TABLA 3.6, Selección de topología para la función de activación Wavelet.

Numero de neuronas en la capa oculta	e_{max}	r	B	E	$T= e_{max} * B * E / r $
1	0.0652	0.435	0.00047421	-7.6944	0.000546895
2	0.0668	0.4743	-0.0008430	-2.1629	0.000256808
3	0.0676	0.4679	-0.0001629	-2.1237	4.99967E-05
4	0.0673	0.4618	-0.0002065	-2.0224	6.08652E-05
5	0.0671	0.4602	-0.0001461	-1.9979	4.2586E-05
6	0.0701	0.5204	-0.0005674	-1.6261	0.000124285
7	0.0752	0.5688	0.00082807	-0.8426	9.22458E-05
8	0.0767	0.5812	0.00090766	-0.7708	9.23283E-05

TABLA 3.7, Selección de topología para la función de activación Sigmoidea.

Numero de neuronas en la capa oculta	e_{max}	r	B	E	$T= e_{max} * B * E / r $
1	0.0664	0.4654	0.0016	-3.3053	0.00075452
2	0.067	0.441	0.00072864	-2.5032	0.00027711
3	0.0658	0.5104	0.00055863	-1.6285	0.00011728
4	0.0693	0.5203	-0.00025269	-1.3617	4.583E-05
5	0.0677	0.5004	0.0008957	-1.5924	0.00019297
6	0.0684	0.5002	0.00084755	-1.4932	0.00017306
7	0.0715	0.5113	0.00016671	-1.0578	2.466E-05
8	0.0692	0.5068	0.001	-1.3468	0.0001839
9	0.0726	0.4972	0.00019381	-1.0401	2.9435E-05
10	0.0696	0.497	0.0012	-1.1209	0.00018837
11	0.0681	0.4739	0.0013	-1.2284	0.00022948
12	0.0678	0.4754	0.0014	-1.2658	0.00025273
13	0.069	0.4897	0.0014	-1.1271	0.00022234
14	0.0721	0.4925	0.00058492	-0.7604	6.5113E-05
15	0.0675	0.4759	0.0013	-0.9283	0.00017117

TABLA 3.8, Selección de topología para la función de activación Tangente hiperbólica.

Numero de neuronas en la capa oculta	e_{max}	r	B	E	$T= e_{max} * B * E / r $
1	0.0689	-0.3281	0.00041042	-11.52232	0.00099307
2	0.0779	0.3634	-0.0013	-5.1717	0.00144122
3	0.0732	0.4368	-0.00013007	-4.5675	9.956E-05
4	0.0666	0.4843	0.00058035	-2.4871	0.00019849
5	0.0665	0.4708	0.00022704	-2.1524	6.9026E-05
6	0.0677	0.4409	0.0016	-2.6568	0.00065272
7	0.067	0.4861	-0.00036751	-1.8446	9.3437E-05
8	0.0667	0.4568	0.00040091	-2.1	0.00012293
9	0.0663	0.449	0.00075947	-2.3357	0.00026194
10	0.0661	0.4593	0.00086297	-2.0927	0.0002599
11	0.0668	0.4043	0.00098092	-2.5063	0.0004062
12	0.0674	0.4435	0.00042836	-2.1398	0.0001393
13	0.0659	0.4545	0.00062428	-2.0005	0.00018108
14	0.0661	0.4517	0.0004199	-2.009	0.00012345
15	0.0657	0.4531	0.00079264	-2.0202	0.00023219

TABLA 3.9, Selección de topología para la función de activación Gaussiana.

Numero de neuronas en la capa oculta	e_{max}	r	B	E	$T= e_{max} * B * E / r $
1	0.0652	0.435	0.00047421	-7.6944	0.00054689
2	0.0668	0.4743	-0.00084304	-2.1629	0.00025681
3	0.0676	0.4679	-0.00016295	-2.1237	4.9997E-05
4	0.0673	0.4618	-0.00020651	-2.0224	6.0865E-05
5	0.0671	0.4602	-0.00014619	-1.9979	4.2586E-05
6	0.0701	0.5204	-0.0005674	-1.6261	0.00012428
7	0.0752	0.5688	0.00082807	-0.8426	9.2246E-05
8	0.0767	0.5812	0.00090766	-0.7708	9.2328E-05

En donde e_{max} es el error máximo en el modelo, r es el coeficiente de correlación, B es el coeficiente Bias y E es el coeficiente de eficiencia. Cabe señalar que con algunas funciones de activación la red neuronal se comporta de manera inestable y es por eso que en algunos casos solo se pudieron generar resultados con hasta ocho neuronas, esto es debido a que el incremento en el aprendizaje es muy rápido y es por esto que en la práctica se utiliza un coeficiente de aprendizaje con el cual se puede realizar un ajuste más fino. Ya que el objetivo de esta sección es determinar cuál es la mejor topología de red neuronal se determino que se debía de manejar un coeficiente de aprendizaje igual para todos los casos. En la figura 3.5 se puede observar el comportamiento de las redes neuronales con función de activación Wavelet, Sigmoidea y Gaussiana. De las cuales se obtuvieron los mejores resultados y siendo la red neuronal con función de activación Gaussiana la que presento mejores resultados se decidió a partir de este punto solo trabajar con ella. También cabe señalar que el modelado mejora mientras se incrementa el número de neuronas hasta llegar a la séptima, esto es porque el coeficiente de aprendizaje, numero de épocas y el numero de neuronas juegan un papel en común el cual encuentra su punto optimo en las siete neuronas pero para el caso del modelado final se utilizaran mas épocas por lo cual se eligió utilizar la red con ocho neuronas en la capa oculta y función de activación Gaussiana.

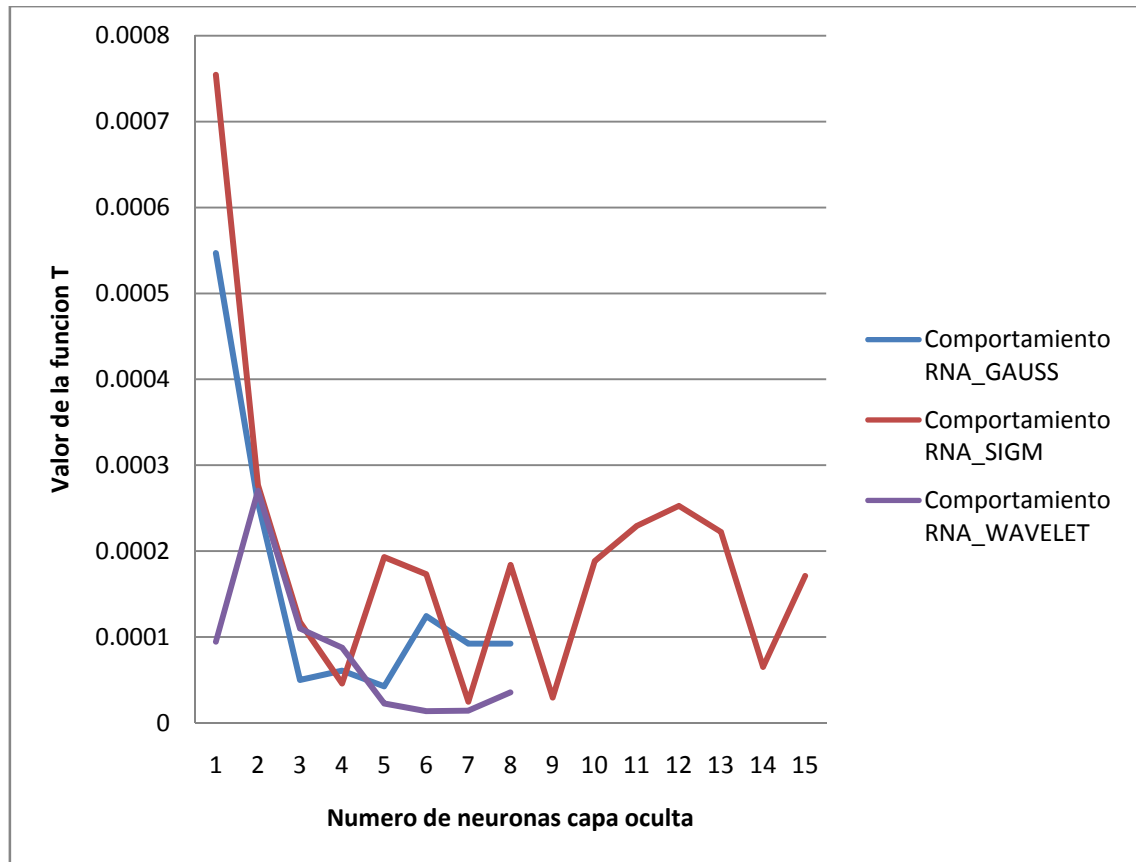


FIGURA 3.5, Comparación de distintas topologías de redes neuronales.

3.6 Diseño del sistema electrónico

El sistema electrónico debe de ser capaz de medir la densidad, conductividad y temperatura que serán las entradas a la red neuronal. Por la parte del procesamiento digital de los datos se decidió implementar la red neuronal en un FPGA ya que con esto podremos trabajar con variables de la longitud que se desee para obtener así la precisión que deseamos. En cuanto al despliegue de datos se utilizara una pantalla LCD y un modulo RF con el cual se enviaron los datos a una base de datos en internet para su posterior análisis.

3.6.1 Medición de densidad

La medición de densidad se realizara por medio de la medición de presión diferencial en una columna de leche como se muestra en la figura 3.6, para realizar la medición de la presión diferencial necesitamos conocer los limites en los cuales variara la presión por lo que a continuación obtendremos estos datos.

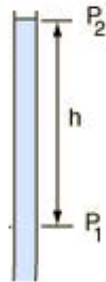


FIGURA 3.6, Medición de densidad por presión diferencial.

Sabemos que el volumen de un cilindro es $V=A \cdot h$, también sabemos que la densidad es $\rho=m/V$, también conocemos que la presión es $P=F/A$ finalmente podemos decir que la fuerza que se tendrá en el sistema será el peso de la leche el cual es $W=m \cdot g$. Entonces podemos obtener la ecuación 71 la cual relaciona la presión con la densidad.

$$\rho = \frac{m}{V} = \frac{\frac{P \cdot A}{g}}{\frac{A \cdot h}{1}} = \frac{P}{g \cdot h} \quad (71)$$

Si se considera una altura constante entonces se puede considerar al denominador de la ecuación 71 como una constante entonces la relación entre la presión y la densidad será lineal. Se considerara a partir de este momento que la altura de la columna será de 10 cm para poder continuar con los cálculos. A partir de la ecuación 71 y conociendo que la altura será de 10 cm y que la densidad de la leche varia en un intervalo de 1015 a 1040 Kg/m³ es posible calcular la variación de presión lo cual se describe en la ecuación 72.

$$\Delta P = P_2 - P_1 = \rho_2 \cdot g \cdot h - \rho_1 \cdot g \cdot h = (\rho_2 - \rho_1) \cdot g \cdot h = 24.53 Pa \quad (72)$$

Ya que la variación de presión es pequeña se decidió utilizar el sensor de presión diferencial con mayor sensibilidad y este es el MPXV7002DP el cual tiene una sensibilidad de 1 V/KPa cabe señalar que no existe en el mercado un sensor con mayor sensibilidad. Esto se compensara con un ADC de 16 bits lo cual nos permitirá tener un tamaño de paso de 76.293μPa. Conociendo el tamaño de paso y la ecuación 71 podemos generar una nueva ecuación que relacione el número de cuenta del ADC con la densidad esta es la ecuación 73. En la figura 3.7 se presenta una imagen del sensor de presión diferencial utilizado.

$$\rho = \frac{196.2}{2^{16}} nADC - 138.7 \quad (73)$$



FIGURA 3.7, Sensor de presión diferencial.

3.6.2 Medición de conductividad

La medición de conductividad puede realizarse directamente por un sensor, en este caso se eligió utilizar el sensor ST3254.2 el cual tiene un rango de medición de 0-100 mS y su salida es de tipo analógica de 4-20mA. Como en el caso anterior podemos generar una ecuación que relacione directamente el número de cuentas del ADC con el valor de la variable que estamos midiendo, en este caso la conductividad, esta es la ecuación 74. En la figura 3.8 se muestra el sensor utilizado para realizar la medición de la variable conductividad.

$$mS = \frac{156.25}{2^{16}} \cdot nADC - 25 \quad (74)$$



FIGURA 3.8, Sensor de conductividad.

3.6.3 Medición de temperatura

Ya que la medición de la temperatura se puede realizar directamente por un sensor, en este caso solo se debe de tener un criterio para elegirlo. Para la medición de temperatura existen cuatro tipos de transductores; termopares, termistores, RTD y semiconductores. Los primeros dos presentan no linealidades que deben de ser compensadas según la aplicación, por otro lado los RTD no presentan no linealidades significativas pero estos son los sensores de temperatura más caros y ya que el rango de temperaturas en los que trabajara el sensor no son muy altos se eligió un sensor de temperatura DS18B20 el cual tiene una sensibilidad igual al de las RTD solo que su rango de operación es menor pero a cambio es te es barato y además tiene integrado un modulo de comunicaciones 1-wire el cual puede ser conectado directamente a los pines del FPGA. En la figura 3.9 se muestra una imagen del sensor de temperatura utilizado.



FIGURA 3.9, Sensor de temperatura.

3.6.4 Diseño del sistema mínimo del FPGA

Ya que este parte del trabajo es bastante amplia se agregara en los anexos un artículo titulado "Diseño de tarjetas electrónicas para dispositivos FPGA" el cual fue escrito durante la maestría y fue publicado en el coloquio de posgrado. En este apartado solo se incluirá una imagen que resume el sistema mínimo con el cual debe de contar un FPGA para que este pueda funcionar de manera adecuada.

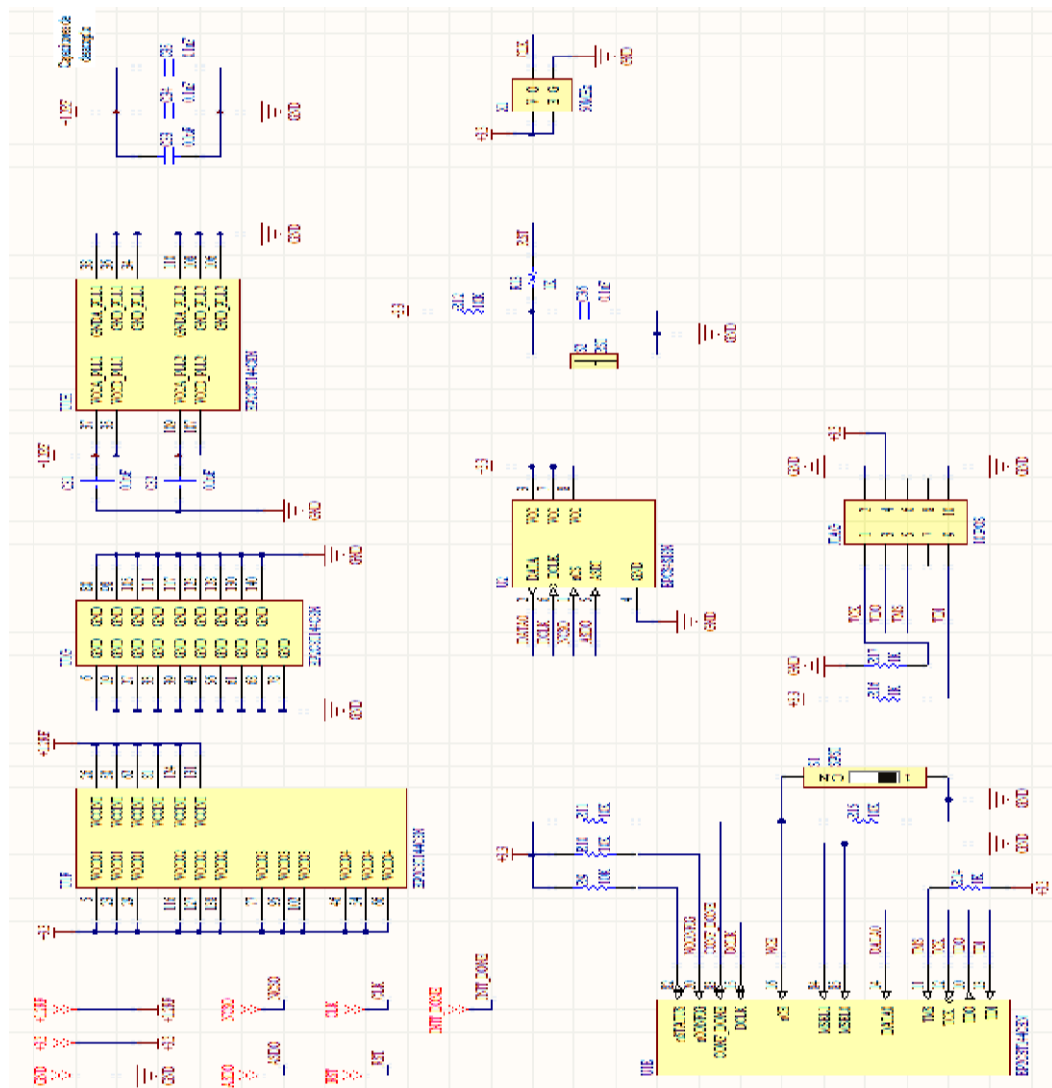


FIGURA 3.10, Configuración de un dispositivo FPGA.

3.6.5 Diseño del PCB del sistema completo

El diseño del sistema se encuentra en los anexos en donde se puede mirar con detalle los esquemáticos así como la lista de componentes, también se agrega el proyecto en formato digital en el CD adjunto a esta tesis. En esta sección solo se mostrar dos imágenes que son el top-view y el bottom-view del PCB diseñado. Figura 3.11 y figura 3.12 correspondientemente.

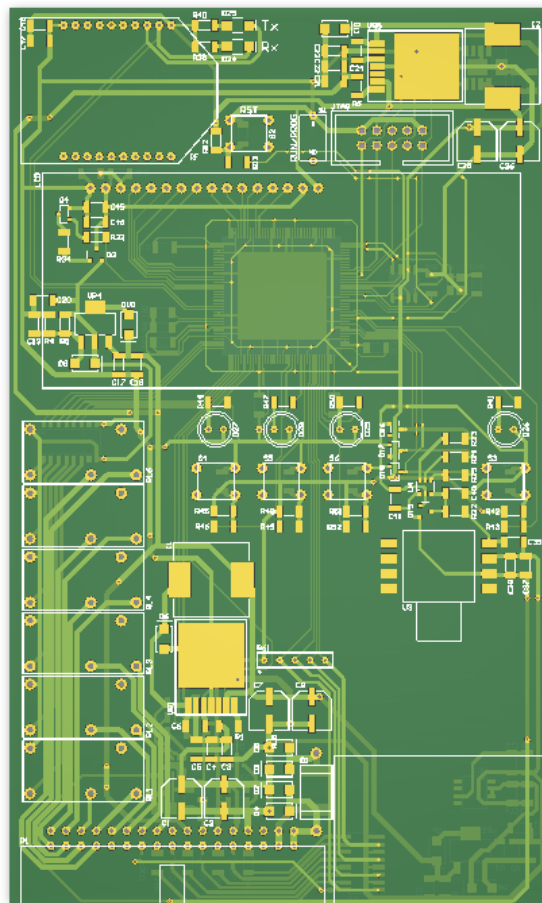


FIGURA 3.11, Top-view del PCB diseñado.

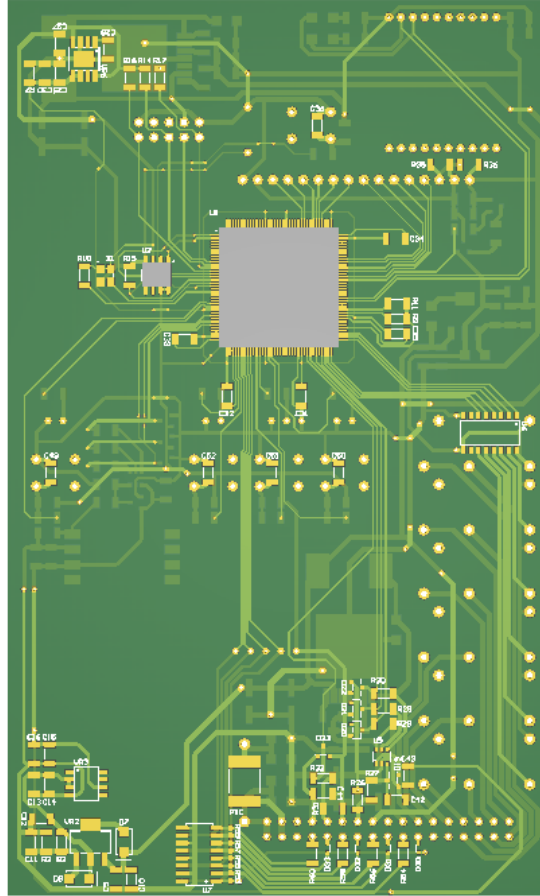


FIGURA 3.12, Bottom-view del PCB diseñado.

3.7 Implementación de la red neuronal en FPGA

El FPGA utilizado es de la marca altera por lo tanto el proyecto se desarrollo en el entorno de trabajo de QUARTUS II. Ya que este apartado es muy extenso se ha decidido describirlo con más detalle en los anexos y de igual forma se agregara el proyecto digital en el CD adjunto a esta tesis. Solo se mencionara que el proyecto consta de 48 archivos .vhd y es por ello que no sería apropiado presentarlo en esta sección.

4. Resultados

En esta sección se presentaran los resultados de la tesis, a grandes rasgos los resultados se pueden dividir en cuatro categorías: los resultados de la fabricación y funcionamiento del prototipo electrónico, los resultados del modelo matemático implementado en MATLAB, los resultados del modelo matemático implementado en FPGA y finalmente los resultados de la prueba de validación.

4.1 Resultados del prototipo electrónico

La tarjeta electrónica funciona según lo previsto, esta cuenta con:

- Se alimenta por medio de una fuente externa 12VCD.
- Fuente interna de 5VCD de precisión.
- Interfaz para sensor de presión diferencial.
- Interfaz para un sensor de 4-20 mA.
- Interfaz 1-Wire con alimentación.
- 6 salidas a relevador con una configuración STSP-NO.
- 4 entradas digitales a 12 VCD.
- FPGA Altera EP2C8T144C8N.
- Memoria de configuración EPCS4SI8N.
- Puerto de programación y consulta JTAG.
- Oscilador de 50 MHz.
- Pantalla LCD 16*2.
- Modulo de RF XBEE.

En la figura 4.1 se muestra el top-view de la tarjeta ensamblada mientras que en la figura 4.2 se muestra el bottom-view de la tarjeta ensamblada.

Finalmente en la figura 4.3 se muestran los recursos que se consumen para poder implementar el sistema digital en el FPGA.



FIGURA 4.1, Top-view de la tarjeta ensamblada.



FIGURA 4.2, Bottom-view de la tarjeta ensamblada.



FIGURA 4.3, Recursos necesarios para implementar el sistema en el FPGA.

4.2 Resultados del modelo implementado en MATLAB

El modelo de la red neuronal final fue implementada en primera instancia en MATLAB dado la facilidad de programación y el despliegue de datos. En la tabla 4.1 se presenta la primera parte de los pesos de interconexión W_{jk} mientras que en la tabla 4.2 se presenta la segunda parte de los pesos de interconexión W_{jk} . Finalmente en la tabla 4.3 se presentan los pesos de interconexión W_{ij} . Estos son los pesos de interconexión del modelo final implementados en MATLAB. Debido a que la red neuronal solo puede representar valores entre 0 y 1 es necesario aplicar una transformación en los datos de entrada y salida. En la figura 4.4 se presenta la variable grasa para la salida normalizada de la red neuronal contra el valor del patrón normalizado. En la figura 4.5 se presenta la variable lactosa para la salida normalizada de la red neuronal contra el valor patrón normalizado. En la

figura 4.6 se presenta la variable proteína para la salida normalizada de la red neuronal contra el valor del patrón normalizado.

TABLA 4.1, Pesos de interconexión W_{ij} parte I.

Entradas	Neurona 1	Neurona 2	Neurona 3	Neurona 4
1	-1.283	2.013	2.686	2.063
2	0.939	1.801	2.185	0.454
3	0.808	0.081	-1.681	0.099

TABLA 4.2, Pesos de interconexión W_{ij} parte II.

Entradas	Neurona 5	Neurona 6	Neurona 7	Neurona 8
1	1.710	2.217	1.835	1.749
2	0.200	0.517	-1.333	0.612
3	-1.001	-1.981	-0.963	-1.093

TABLA 4.3, Pesos de interconexión W_{jk} .

Neurona CO	Neurona SAL 1	Neurona SAL 2	Neurona SAL 3
1	0.773	-1.649	1.025
2	2.506	-0.117	0.301
3	2.223	-0.423	0.085
4	0.533	0.893	1.325
5	-2.138	1.430	-0.051
6	4.125	0.825	0.531
7	-1.625	-1.462	-0.918
8	-2.149	1.730	0.259

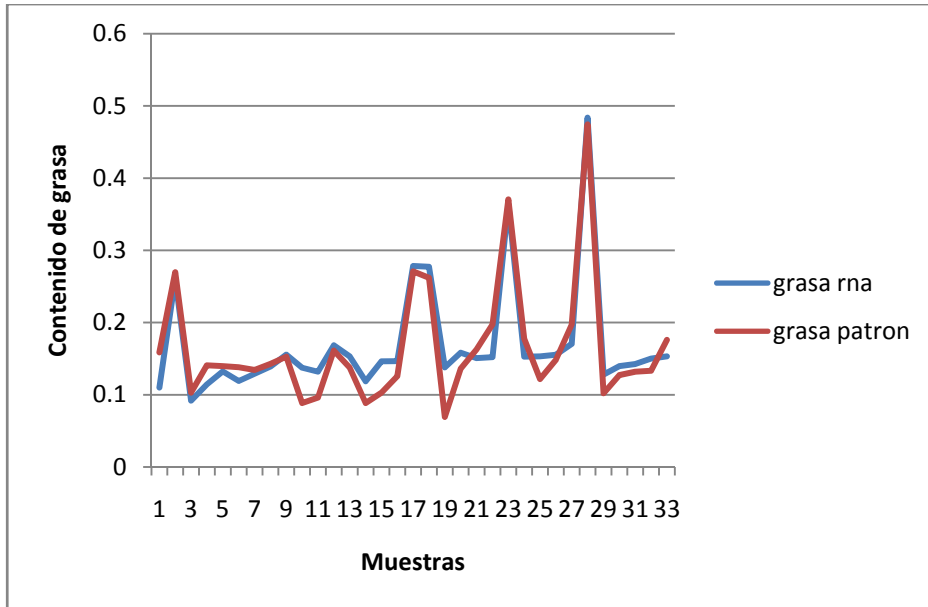


FIGURA 4.4, Comparación variable grasa MATLAB.

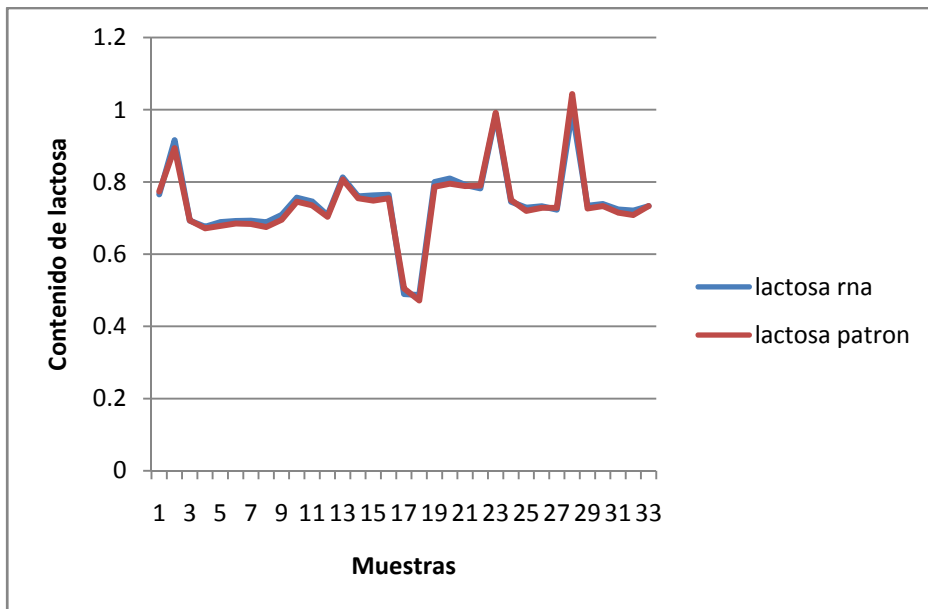


FIGURA 4.5, Comparación variable lactosa MATLAB.

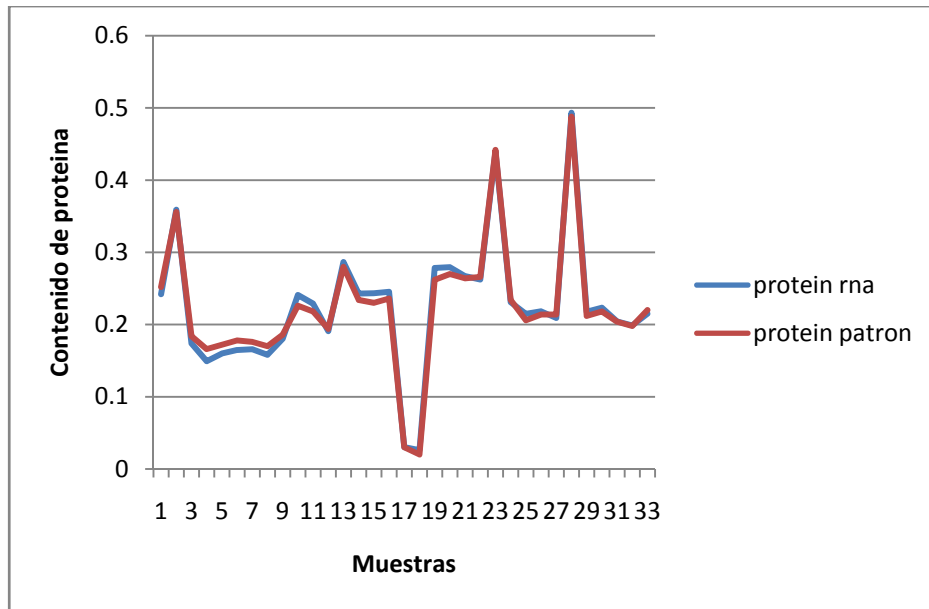


FIGURA 4.6, Comparación variable proteína MATLAB.

En la figura 4.7 se puede observar el porcentaje de exactitud por muestra para las tres variables de salida de la red neuronal implementada en MATLAB. Se puede observar que la variable que presenta mas error es el contenido de grasa, sin embargo la exactitud de esta variable se mantiene en un rango de 97.5% a 102.5% por lo que el error se mantiene en el rango del 5% que es lo que se deseaba mediante el uso de redes neuronales.

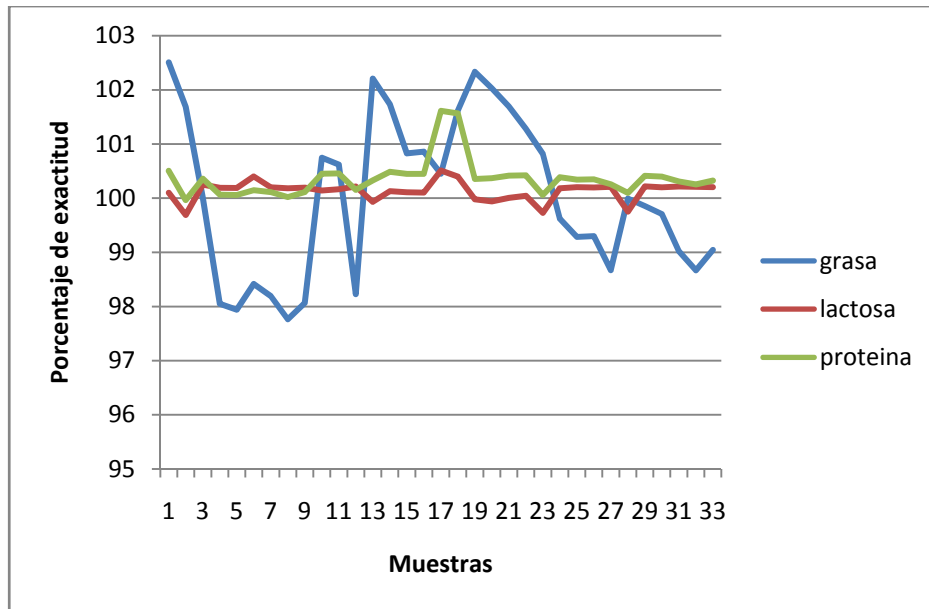


FIGURA 4.7, Porcentaje de exactitud del modelo en MATLAB.

4.3 Resultados del modelo implementado en FPGA

Los pesos de interconexión de la red neuronal se implementaron con un formato 6.26 con signo. En la tabla 4.4 se presenta la primer parte de los pesos de interconexión W_{ij} , en la tabla 4.5 se presenta la segunda parte de los pesos de interconexión W_{ij} y finalmente en la tabla 4.6 se presentan los pesos de interconexión W_{jk} .

TABLA 4.4, Pesos de interconexión W_{ij} parte I en formato 6.26.

Entradas	Neurona 1	Neurona 2	Neurona 3	Neurona 4
1	FFFADD6680	80DED6A	ABF3BA4	8415A18
2	3C195AD	7348127	8BE152F	1D11694
3	33BA81B	539B1F	FFF945CD87	665AA5

TABLA 4.5, Pesos de interconexión W_{ij} parte II formato 6.26.

Entradas	Neurona 5	Neurona 6	Neurona 7	Neurona 8
1	6D76453	8DECF81	757F279	6FF8D06
2	CD988F	2122806	FFFAAA9E5B	272C13E
3	FFFBFEAD86	FFF81312C2	FFFC25018E	FFFBA09C4D

TABLA 4.6, Pesos de interconexión W_{jk} formato 6.26.

Neurona CO	Neurona SAL 1	Neurona SAL 2	Neurona SAL 3
1	317D2FA	FFF96735EC	41A493D
2	A06FEC6	FFFF87C623	134C23C
3	8E49F7E	FFFE4E1045	5736B4
4	2229196	3934CA7	54D187C
5	FFF7728937	5B8A8CF	FFFBCB33E5
6	1080C189	34D9B6B	220B1CA
7	FFF97F6611	FFFA2691A3	FFFC536E4E
8	FFF766E1A0	6EC0D5C	10A26EA

En la figura 4.8 se presenta la variable grasa para la salida normalizada de la red neuronal en FPGA contra el valor normalizado del instrumento patrón. En la figura 4.9 se presenta la variable lactosa para la salida normalizada de la red neuronal en FPGA contra el valor normalizado del instrumento patrón. En la figura 4.10 se presenta la variable proteína para la salida normalizada de la red neuronal en FPGA contra el valor normalizado del instrumento patrón. Finalmente en la figura 4.11 se observa el porcentaje de exactitud por muestra el cual no difiere significativamente con los resultados obtenidos por la implementación en MATLAB.

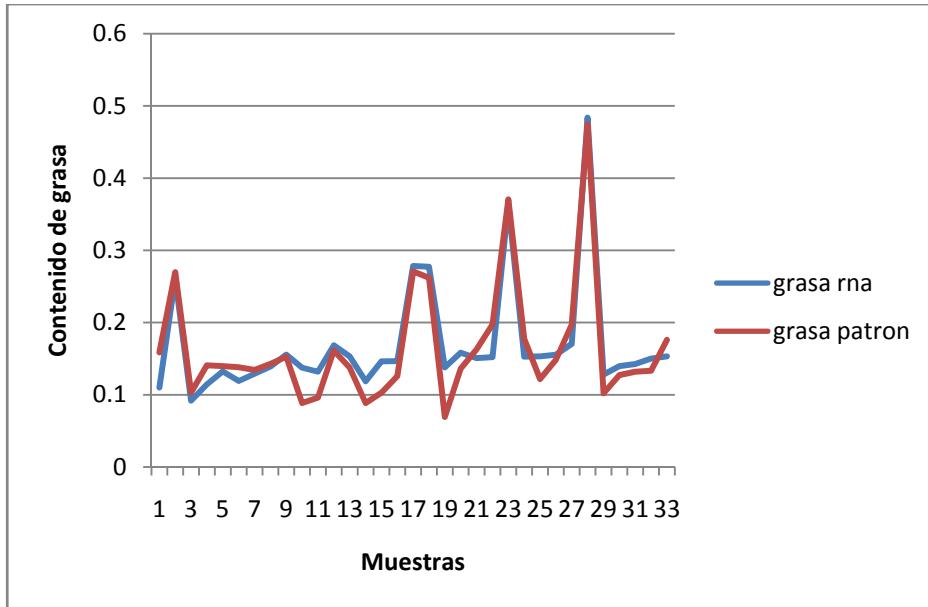


FIGURA 4.8, Comparación variable grasa FPGA.

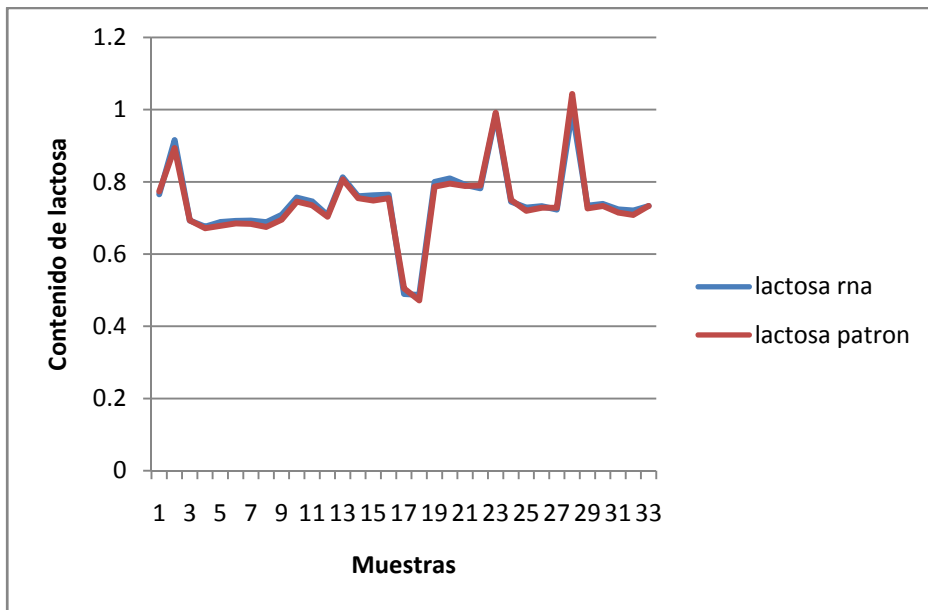


FIGURA 4.9, Comparación variable lactosa FPGA.

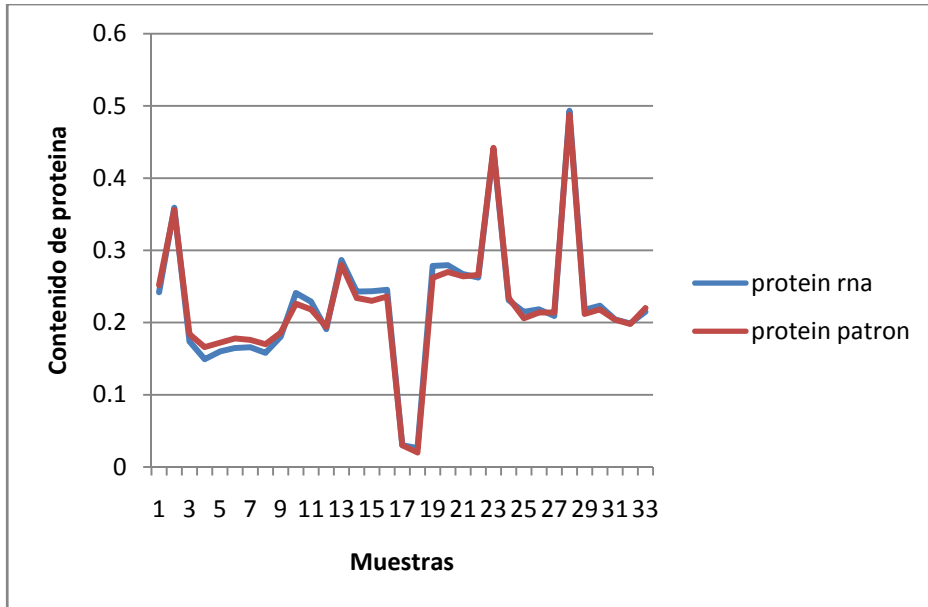


FIGURA 4.10, Comparación variable proteínas FPGA.

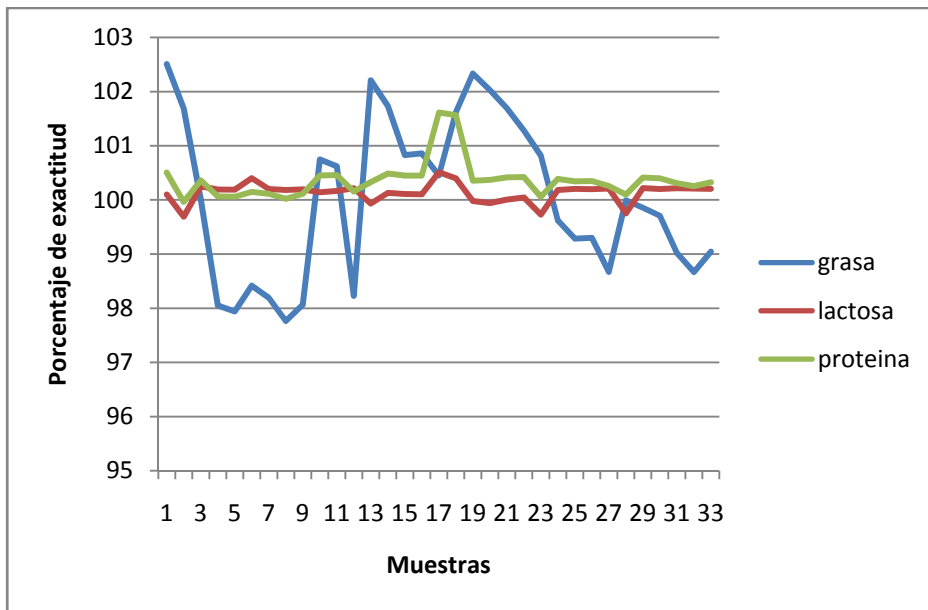


FIGURA 4.11, Comparación exactitud por muestra FPGA.

5. Conclusiones

Los resultados obtenidos en este trabajo claramente son los esperados ya que para la variable grasa se obtuvo un coeficiente de correlación $r=0.9505$ un coeficiente Bias $B=-0.0048$ y un coeficiente de eficiencia $E=0.8877$. Para la variable lactosa se obtuvo un coeficiente de correlación $r=0.9936$ un coeficiente Bias $B=-0.0036$ y un coeficiente de eficiencia $E=0.9848$. Para la variable proteína se obtuvo un coeficiente de correlación $r=0.9944$ un coeficiente Bias $B=-0.000593$ y un coeficiente de eficiencia $E=0.9887$.

Estos resultados son completamente los esperados ya que en el diseño experimental se busco obtener un nivel de confianza del 95% lo cual se ve reflejado en el coeficiente de correlación de la variable grasa. Por otro lado se confirmo que el coeficiente de correlación de la variable grasa seria más bajo que el de la variable lactosa y siendo la variable proteína la que presenta el mejor modelo, lo cual se predijo desde que se realizo la prueba de normalidad de Anderson Darling. Esto se debe a que esta variable presenta un mayor nivel de confianza con una distribución normal.

La función T propuesta en este trabajo es de gran utilidad ya que provee de información con la cual se puede tomar decisiones fundamentadas en hechos, también cabe mencionar que una buena modificación a esta función agregaría el numero de neuronas así como algún tipo de medida que involucre la velocidad de convergencia de cada red en específico.

6. Referencias

Moctezuma, L. G., J. A. Espinosa, R. V. Cuevas, B. J. Jolalpa y S. F. Romero. 2008. Priorización de Cadenas Agroalimentarias como Herramienta de Toma de Decisiones en la Región Centro de México: Casos Hidalgo y Querétaro. Memoria in extenso del XXI Congreso Internacional en Administración de Empresas Agropecuarias. SOMEXAA. Torreón, Coah. pp. 996-1007.

- Cervantes, F. E., A. V. Cesín y I. O. Mamani. 2012. La calidad estándar de la leche en el estado de Hidalgo. Revista Mexicana de Ciencias Pecuarias. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias.
- Pinto, M. C., E. R. Carrasco, B. L. Fraser, A. H. Letelier y W. P. Dorner. 1998. Composición química de la leche cruda y sus variaciones a nivel de silos en plantas lecheras de la VIII, IX y X regiones de Chile. Revistas electrónicas UACH.
- CANILEC. 2011a. Apéndice informativo B: Sustancias nutritivas. Tercera sección: Secretaria de salud.
- CANILEC. 2011b. Apéndice informativo C: Métodos de prueba. Tercera sección: Secretaria de salud.
- FIL-IDF. 1983. Nutrition and metabolism. Report at IDF Annual Sessions 1982. FIL-IDF. Bull 166./1983. Brussels, Belgium.
- FIL-IDF. 1995. Milk payment systems for ex-farm milk. FIL-IDF; Bulletin N° 305, 51 pp. Brussels, Belgium.
- Bruckmaier, R. M.. 2001. Milk ejection during machine milking in dairy cows. Livestock Production Science. Publ. No. 70 p 121-124.
- Bi, GQ, M. M. Poo. 2001. Synaptic modification by correlated activity: Hebb's postulate revisited. Annu Rev Neurosci 24:139–166.
- Burkitt, AN, H . Meffin, D. B. Grayden. 2004. Spike-timing-dependent plasticity: The relationship to rate-based learning for models with weight dynamics determined by a stable fixed point. Neural Comput 16(5):885–940.
- Calderón, A., F. García y G. Martínez. 2006. Indicadores de calidad de leches crudas en diferentes regiones de Colombia. Revista MVZ Córdoba. Volumen 11 (1).
- CANILEC. 2011a. Apéndice informativo B: Sustancias nutritivas. Tercera sección: Secretaria de salud.
- CANILEC. 2011b. Apéndice informativo C: Métodos de prueba. Tercera sección: Secretaria de salud.
- Castro, L. C., G. Sánchez, L. F. Iruegas y G. Saucedo. 2001. Tendencias y oportunidades de desarrollo de la red leche en México. FIRA Boletín Informativo XXXIII (317). México.
- Cervantes, F. E., A. V. Cesín y I. O. Mamani. 2012. La calidad estándar de la leche en el estado de Hidalgo. Revista Mexicana de Ciencias Pecuarias. Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias.

- Funahashi, K. 1989. On the approximate realization of continuous mapping by neural networks. *Neural Networks*, No. 2, pp. 183-192.
- Gilson, M., A. N. Burkitt, D. B. Grayden, D. A. Thomas y J. L. Hemmen. 2009. Emergence of network structure due to spike-timing-dependent plasticity in recurrent neuronal networks. I. Input selectivity-strengthening correlated input pathways. *Biological Cybernetics*. 101:81-102.
- Gütig, R., R. Aharonov, S. Rotter, H. Sompolinsky. 2003. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. *J Neurosci* 23(9):3697–3714.
- Hornik, K. 1989. Feedforward networks are universal approximators. *Neural Networks*, No. 2, pp. 359-366.
- Markram, H, J. Lubke, M. Frotscher, A. Roth, B. Sakmann. 1997. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *J Physiol (Lond)* 500(2):409–440.
- Meffin, H, J. Besson, A. N. Burkitt, D. B. Grayden. 2006. Learning the structure of correlated synaptic subgroups using stable and competitive spike-timing-dependent plasticity. *Phys Rev E* 73(4):041911.
- Moctezuma, L. G., J. A. Espinosa, R. V. Cuevas, B. J. Jolalpa y S. F. Romero. 2008. Priorización de Cadenas Agroalimentarias como Herramienta de Toma de Decisiones en la Región Centro de México: Casos Hidalgo y Querétaro. Memoria in extenso del XXI Congreso Internacional en Administración de Empresas Agropecuarias. SOMEXAA. Torreón, Coah. pp. 996-1007.
- Narendra, K.S. and K. Parthasaraty. 1990. Identification and Control of Dynamical Systems using Neural Networks, *IEEE Transactions on Neural Networks*, Vol. 1, No. 1, pp. 5-27.
- Pineda, F. J. 1989. Recurrent back propagation and the dynamical approach to adaptive neural computation. *Neural Computation* 1, pp. 161-172.
- Pinto, M. C., E. R. Carrasco, B. L. Fraser, A. H. Letelier y W. P. Dorner. 1998. Composición química de la leche cruda y sus variaciones a nivel de silos en plantas lecheras de la VIII, IX y X regiones de Chile. *Revistas electrónicas UACH*.
- Pizarro, C., I. Esteban-Díez, J. M. González-Sáiz y M. Forina. 2007. *J. Agric. Food Chem.* 55, 7477-7488.
- Vázquez-Valencia, R. A. y I. Aguilar-Benítez. 2010. Organizaciones lecheras en los altos sur de Jalisco: un análisis de las interacciones productivas. *Región y Sociedad*. 1870-3925.

- Zorrilla, R. J., V F. Ramírez, G. G. Salazar, L. R. Martínez y R. F. J. Padilla. 1997. Descripción cualitativa y cuantitativa del sistema de lechería familiar en los Altos de Jalisco, Instituto Nacional de Investigaciones Forestales–Centro para la Investigación en Sistemas Sostenibles de Producción Agropecuaria (INIFAP–CIPAV). Memorias del Seminario–taller en sistemas de producción de leche en pequeña escala. Universidad Autónoma del Estado de México, Jalisco.
- Zotte, A. D., P. Berzaghi, L. M. Jansson y I. Andrighetto. 2006. *Animal Feed Sci. Technol.* 128, 108-121.