

Generación de un modelo mediante el uso de redes neuronales artificiales para la detección de mastitis en vacas lecheras del INTA Estación Experimental Agropecuaria Rafaela

Mariano Ferrero
Universidad Católica de Santiago del Estero
(Departamento Académico Rafaela)
mariano-ferrero@live.com.ar

Resumen. El presente informe corresponde al Trabajo Final de Grado de la carrera Ingeniería en Informática de la Universidad Católica de Santiago del Estero, Departamento Académico Rafaela. El objetivo del trabajo era facilitar la detección de mastitis en vacas lecheras del Instituto Nacional de Tecnología Agropecuaria Estación Experimental Agropecuaria Rafaela, utilizando herramientas de software, más específicamente, redes neuronales artificiales.

La presencia de mastitis en vacas lecheras es uno de los eventos sanitarios con mayor incidencia en los rodeos de la Argentina. La enfermedad tiene como principal característica, luego de un período de tiempo, la disminución en los volúmenes de producción, causando pérdidas económicas.

El trabajo incluye una descripción y justificación del problema abordado, desarrollo del marco teórico, marco metodológico y las distintas fases que permitieron construir y entrenar diversos modelos de redes neuronales.

En la etapa final se realizó dicho entrenamiento con distintas redes del tipo Perceptrón Multicapa, alcanzando diversos resultados. El mejor rendimiento se obtuvo a partir de la selección de aquellos animales que se encontraban en su primer tercio de lactancia, utilizando 25 neuronas en la capa oculta y una segmentación de los conjuntos de estimación, validación y prueba de 80%, 10% y 10% respectivamente.

Palabras claves: inteligencia artificial • redes neuronales • mastitis • perceptrón multicapa.

1. Justificación del proyecto

La presencia de mastitis en vacas lecheras tiene como consecuencia la disminución de los volúmenes de producción obtenidos sobre el animal afectado.

En la Fig. 1 se introduce un gráfico que contiene la representación de la lactación de una misma vaca para dos períodos distintos, uno donde contrajo mastitis clínica a los 60 días (serie de color verde) y otro donde no lo hizo (serie de color rojo).

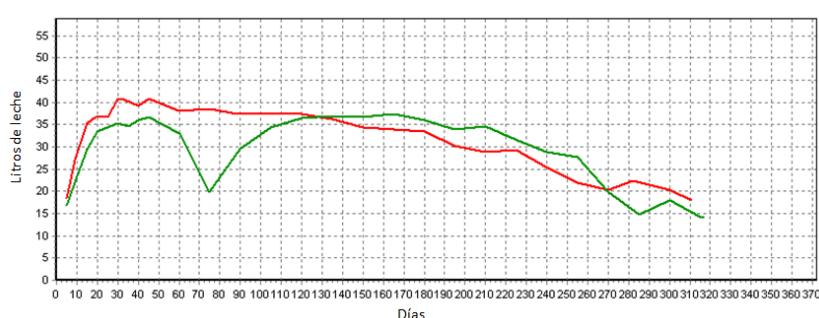


Fig. 1. Gráfico de lactaciones obtenido del aplicativo ALPRO Windows

Como se puede apreciar, existe una divergencia notoria en la producción de leche de uno y otro caso. De acuerdo a lo expresado en un artículo del periódico MOTIVAR en su edición de julio de 2013, bajo el nombre de “Un tambo promedio pierde 35 días de producción a causa de mastitis” [1], se nombran estadísticas obtenidas sobre esta temática en la Argentina, como por ejemplo, la pérdida anual de producción en el país es equivalente a lo consumido por 1.300.000 argentinos a causa de esta enfermedad. A su vez, tal como menciona el nombre del artículo, se señala que un tambo argentino promedio pierde un mes de facturación al año (35 días) por la presencia de mastitis.

Estos datos introducen la idea de que la incidencia y el impacto de esta enfermedad en el país no debieran ser, al menos desde el punto de vista económico, pasados por alto.

2. Estado del arte

Durante la realización del presente estudio se han encontrado diversos trabajos investigativos que relacionan de distintas maneras las temáticas de redes neuronales artificiales y mastitis en vacas lecheras. Es necesario destacar que la gran mayoría de

estos trabajos han sido desarrollados en países de Europa, así como también Nueva Zelanda y Estados Unidos. A su vez, los mismos utilizan información proporcionada por sistemas de ordeño altamente automatizados. En este contexto, a diferencia de lo que ocurre en Argentina donde los animales se encuentran permanentemente en condiciones de pastoreo, los factores ambientales tienen una incidencia prácticamente nula.

Dos de los estudios que se han obtenido se mencionan a continuación.

Detección de mastitis clínica en sistemas lecheros automatizados [2]

El principal objetivo de este artículo fue mejorar la detección de mastitis clínica recurriendo a la evaluación de diversas técnicas, dentro de las cuales se pueden mencionar la utilización de lógica difusa, redes neuronales artificiales, filtros de Kalman sobre series de tiempo, análisis discriminante, entre otras.

Los resultados obtenidos para el uso de redes neuronales artificiales específicamente arrojaron en su mejor rendimiento un valor de sensibilidad igual a 84%, así como también una especificidad de 97%.

Detección de mastitis y su estado de progreso en sistemas lecheros automatizados utilizando redes neuronales artificiales [3]

Dos tipos de redes neuronales artificiales, Perceptrón Multicapa (MLP por sus siglas en inglés) y Self-Organizing feature Map (SOM), fueron utilizados para detectar mastitis en sistemas de ordeño automatizados. Cuatro modelos MLP's con cuatro combinaciones en las variables de entrada fueron desarrollados a fin de detectar cuartos infectados. Como resultado se obtuvo un modelo que permite realizar una clasificación con un rango de aciertos de 90,74%, mientras que los valores de sensibilidad y especificidad fueron de 86,9% y 91,36% respectivamente.

2.1. Marco teórico

Ciclo de lactancia: Se entiende al ciclo de lactancia como “la producción de leche que se obtiene de una vaca en el período de lactancia en un año y está determinado por el intervalo entre partos”, de acuerdo a la definición de Álvaro Castro Ramírez [4].

Inteligencia artificial: Se define como inteligencia artificial a “la automatización de actividades que asociamos con el pensamiento humano, tales como toma de decisión, resolución de problemas o aprendizaje” [5].

Mastitis: Se da conocimiento de este término a “una enfermedad infecto-contagiosa de la glándula mamaria, en la cual la inflamación se produce como respuesta a la invasión, a través del canal del pezón, de diferentes tipos de bacterias, mico plasmas, hongos, levaduras y hasta algunos virus. (...) Clásicamente se la ha definido como una “enfermedad poli factorial”, porque el riesgo de infección depende de la habilidad de la vaca para rechazarla, del tipo, número y patogenicidad de las bacterias presentes en un rodeo y, fundamentalmente, de las condiciones de medio ambiente y del manejo del ordeño en particular que se estén desarrollando en un establecimiento” [6].

Red neuronal: Se entiende como red neuronal a todo “procesador masivo, distribuido y paralelo formado de unidades simples de procesamiento, que posee una natural propensión de almacenar conocimiento experimental y disponerlo para su uso” [7].

3. Marco metodológico

3.1. Tipo de investigación

Para la realización de este trabajo se utilizó un tipo de investigación de campo ya que se desarrolló con ánimos de interpretar y solucionar un problema en un momento determinado.

Los datos utilizados fueron proporcionados por la Estación Experimental Agropecuaria Rafaela del Instituto Nacional de Tecnología Agropecuaria. Dicho vínculo se estableció mediante el convenio que dicha institución y la Universidad Católica de Santiago del Estero poseen, en pos de fomentar los trabajos con características similares al presente.

Por otra parte, acorde a los objetivos intrínsecos o el nivel de conocimiento [8], la investigación ha iniciado como de tipo descriptiva. Esto se fundamenta en que la motivación primordial radicó en describir algunas características fundamentales de conjuntos homogéneos de fenómenos. Asimismo, este tipo de investigaciones utilizan criterios sistemáticos que permiten poner de manifiesto la estructura o el

comportamiento de dichos fenómenos, como ejemplifican los modelos matemáticos de curvas de lactancia ya desarrollados, modelo de Wood [9] por ejemplo.

De acuerdo a los objetivos extrínsecos o externos, se puede clasificar a la investigación como aplicada, ya que ha perseguido objetivos de aplicación práctica en un futuro relativamente cercano.

A su vez, y como último criterio de clasificación para el tipo de investigación, se puede catalogar a la misma como cuantitativa (se utilizaron procedimientos y técnicas estadísticas al utilizar redes neuronales artificiales, ha sido orientada al resultado y, bajo un conjunto de características, generalizable) y longitudinal (se realizó un seguimiento de un determinado conjunto de variables de los mismos animales a través del tiempo).

3.2. Población y muestra

Acorde a la problemática que ha motivado la investigación, la población estuvo determinada por las mediciones realizadas por el INTA EEA Rafaela sobre la producción de leche, así como también registros de incidencias sanitarias y datos meteorológicos para cada fecha en cuestión.

De dicha población se obtuvo una muestra representativa e intencional, con el objetivo de que los resultados obtenidos puedan ser generalizables a la totalidad de los mismos y a su vez se trabaje con una proporción suficiente de información de animales afectados por la enfermedad. El lapso de tiempo seleccionado estuvo comprendido entre el 1 de enero de 2010 al 31 de diciembre de 2012, de acuerdo a la disponibilidad de datos al momento de realizar la investigación.

3.3. Recolección de información

La información a utilizar ha sido obtenida mediante el uso de un sistema automatizado de control de producción lechera. A su vez, los datos son administrados por una aplicación conocida como ALPRO, desarrollado por la empresa DeLaval. Mediante esta aplicación se han realizado sucesivas exportaciones de los datos de interés, los cuales han sido traspasados luego a hojas de cálculo y posteriormente importados a una base de datos SQL Server.

Resulta pertinente aclarar que la técnica de recolección de información fue la utilización de datos secundarios. Dicha clasificación se basa en que los datos han sido previamente recogidos por otros investigadores y son lo suficientemente válidos, pertinentes y confiables como para ser utilizados en este estudio, no siendo necesario de esta forma repetir los procedimientos ya efectuados.

A su vez, los datos correspondientes a incidencias sanitarias fueron obtenidos desde los registros existentes en el laboratorio de la institución previamente nombrada.

Por último, la información meteorológica fue proporcionada por el personal de la entidad, la cual se dispone en almacenamientos internos de la misma. Al igual que los datos sanitarios, estos últimos fueron entregados en archivos con formato de hoja de cálculo, y posteriormente importados a la misma base de datos donde se encontraban los registros de producción lechera.

3.4. Tratamiento de los datos

Luego de obtener la información se prosiguió con el desarrollo de un modelo de datos que permitió almacenarla seleccionando únicamente los datos de interés y adicionando aquellos que admitían representar las variables elegidas. En relación a los datos que presentaban valores irrisorios o nulos, los mismos fueron filtrados a modo de quitar el ruido que podrían insertar en el análisis.

4. Desarrollo

4.1. Elección del modelo de red neuronal a utilizar

El modelo seleccionado para utilizar en este problema ha sido una red Perceptrón Multicapa (MLP), con una capa oculta y una única neurona de salida. Esta elección se basa en que estos modelos son ampliamente utilizados en generalizaciones, es decir, situaciones que requieren relacionar entradas con salidas similares entre sí, de acuerdo a lo desarrollado por Haykin [7].

Es necesario aclarar, a su vez, que este modelo ha sido ampliamente utilizado en problemas similares al que se presenta en este caso [3] [10] [11], de acuerdo a lo obtenido en el proceso de revisión del estado del arte, obteniendo resultados positivos prácticamente en la totalidad de los mismos.

En lo que respecta al número de neuronas utilizado en la capa oculta, se ha realizado un conjunto de pruebas determinado con cantidades diversas y crecientes, comenzando desde 5, a modo de decretar qué configuración ofrecía el mejor rendimiento y, a su vez, permitía una correcta generalización en lugar de meramente memorizar los datos proporcionados.

Debido a que cada neurona de la red utiliza una función de activación no lineal, diferenciable en cada punto pero que a su vez presenta gran fluidez [7], la función que ha sido testeada es la sigmoidea o logística.

Por último, la salida de la red estará representada por un único valor indicando la presencia (valores cercanos a 1) o ausencia (valores cercanos a 0) de la enfermedad en un animal para una fecha determinada, de acuerdo al conjunto de datos ingresado. Dicho conjunto se compone de un único registro por cada una de las variables que se describen en la siguiente sección.

4.2. Análisis y determinación de variables de interés

Dado que la red neuronal debe nutrirse de un conjunto de datos determinado, ha sido necesario realizar una selección de aquellas variables que sean representativas sobre el objeto de estudio, y permitan alcanzar las metas planteadas.

Para justificar la selección de estas variables es necesario aclarar que la enfermedad con la que se ha tratado es considerada multifactorial, y desde el punto de vista epidemiológico se tienen en cuenta tres elementos que componen lo que se conoce como triángulo epidemiológico de la mastitis [12]. Dichos elementos son la ubre del animal, los patógenos y el medio ambiente.

Esta relación se presenta debido a que la enfermedad es el resultado de diversas interacciones de manejo y del medio ambiente, las cuales determinan un aumento en la exposición a organismos causales y, como consecuencia, que se favorezca el paso de estos agentes por el canal del pezón [12]. En adición a esto, Calvinho [13] menciona que las incidencias por patógenos ambientales constituyen un problema significativo en países donde las vacas son mantenidas el 100% del tiempo en condiciones de pastoreo, como por ejemplo Argentina (en la gran mayoría de los casos). Debido a esto serán considerados, además de los datos de producción e incidencias sanitarias, los registros provenientes de meteorología, a fin de contemplar

las condiciones del medio ambiente que podrían favorecer la aparición de la enfermedad.

A continuación se detallarán cada una de las variables que han sido seleccionadas para constituir las entradas al modelo.

Litros de leche: Se conoce que altos volúmenes de producción disminuyen la resistencia de la vaca a la enfermedad.

Porcentaje de variación de hoy con respecto al día previo: La presencia de mastitis presupone un descenso en la producción.

Porcentaje de variación de hoy con respecto al promedio de los últimos 3 días: Esta variable amplía la base histórica (período de mastitis sub-clínica) teniendo presente lo nombrado en la descripción de la variable previa.

Porcentaje de variación de hoy con respecto al promedio del rodeo: Esta variable persigue incluir aquellas variaciones que afectan a la totalidad del rodeo y que no necesariamente determinan la aparición de la enfermedad en un animal puntualmente.

Número de lactación: La producción aumenta progresivamente de un parto con respecto a otro hasta la 5ta lactación, donde comienza a decrecer o se mantiene constante hasta el año 11° [14]. También se considera que la mayor cantidad de casos de mastitis sub-clínicas se encuentran en animales con un avanzado número de lactancias, mientras que los casos clínicos se hacen presente mayoritariamente en vacas con un menor número de lactancias [13].

Días en lactación: Se conoce que la máxima producción se encuentra entre los 40 y 65 días posparto (periodo con mayor incidencia de infecciones intramamarias [13]), donde luego comienza a decrecer levemente hasta los 305 días aproximadamente.

Estación de parto: De acuerdo a lo expresado por Cuatrin [15] la producción es mayor en los partos de primavera respecto a los de otoño, aunque la persistencia de los de otoño sea más elevada.

Incidencias de mastitis en los últimos 365 días: Aquellos casos donde la enfermedad se ha presentado de forma recurrente en el último año para un mismo animal podría significar un caso de mastitis crónica, aumentando considerablemente la posibilidad de ser infectada nuevamente.

Precipitaciones: Las precipitaciones generan en algunos casos situaciones que son desfavorables para la producción de leche debido a que, por ejemplo, el animal no dispone de su alimento en las mismas condiciones que un día seco.

Promedio de precipitaciones en los últimos 10 días: Dado que la incidencia de lluvias en el suelo podría generar la acumulación de barro, se estaría contribuyendo a la aparición de mastitis por organismos coliformes, cuyo principal reservorio es el medio ambiente de la vaca y contaminan la piel de los pezones principalmente en el intervalo entre ordeños [13].

Estrés calórico: Mediante esta variable se contemplan las condiciones de temperatura y humedad presentadas en el contexto del animal. El cálculo de la misma se presenta en la sección *Apéndice 1: Cálculo de estrés calórico*.

Promedio de estrés calórico en los últimos 10 días: Debido a lo nombrado en la descripción de la variable precedente se desea analizar lo acontecido en los días previos al análisis, de modo que pueda contemplarse el aumento en las probabilidades de aparición de la enfermedad al cabo de unos días de presentada dicha situación de estrés calórico en el animal, así como también relacionar un posible descenso en la producción en base a esto.

4.3. Selección de conjuntos de aprendizaje y test

En lo que refiere al propio entrenamiento de la red, se dividió el cúmulo de los datos en conjuntos de estimación, validación y prueba. El primero de estos conjuntos fue utilizado como entrenamiento, mientras que el segundo de ellos se estipuló como validación cruzada de dicho entrenamiento [7]. El set de datos remanente cumplió la función de estimar el comportamiento de la red.

Vale la pena aclarar que en todos los casos los conjuntos estuvieron compuestos por la misma cantidad de casos positivos que negativos. Los porcentajes de asignación

para cada uno de los conjuntos nombrados iteraron entre diferentes valores, a fin de encontrar el mejor rendimiento posible para la red.

4.4. Pre-procesamiento

En esta etapa el objetivo perseguido fue maximizar la eficiencia durante el entrenamiento en relación al conjunto de datos utilizado. Con respecto a esta última actividad, el algoritmo seleccionado que se aplicó fue Back-Propagation (BP), siendo éste el algoritmo estándar a utilizar en redes neuronales del tipo MLP de acuerdo a lo expresado por Haykin [7].

Una vez seleccionado el modo de entrenamiento a llevar a cabo resulta pertinente utilizar heurísticas sugeridas por dicho autor para mejorar el rendimiento del algoritmo, como por ejemplo, la normalización de las entradas con el objetivo de que cada variable tenga una media igual a 0 y una varianza igual a 1, o dicho de otro modo, que cada una de ellas tenga la misma importancia (similar variabilidad y espectro de valores). Para llevar esto a cabo las variables con valores discretos fueron binarizadas, mientras que las restantes han sido normalizadas en base a la Ecuación 1.

$$x' = \frac{x - \tilde{x}}{\sigma}$$

Ecuación 1. Normalización para las variables de entrada.

A su vez, el rango para estas últimas variables se estipuló entre los valores -1 y 1, de acuerdo a la Ecuación 2.

$$x'' = \frac{2 * (x'_j - \min x')}{\max x' - \min x'}$$

Ecuación 2. Fijación del rango de valores para las entradas.

Por otro lado, la maximización del contenido que los datos ofrecen es esencial en el desempeño de esta etapa. Para esto se especifican dos actividades fundamentales [7]:

- Utilizar un dato de ejemplo que arroje el mayor error posible de entrenamiento.
- Cada ejemplo utilizado debe ser radicalmente diferente a todos los proporcionados anteriormente en ese conjunto.

Esto ha sido afrontado utilizando un algoritmo que seleccione de forma aleatoria los valores de entrada para cada iteración. A su vez, dado que la cantidad de registros que

indican la presencia de mastitis es significativamente menor a aquellos casos en los que se trata de una vaca sana, se repitió la fase de entrenamiento la cantidad de veces necesaria hasta que todos los datos sean contemplados por igual, tomando en cada caso la misma cantidad de positivos y negativos.

4.5. Entrenamiento

La automatización de este proceso contó con las siguientes características:

- Se utilizó MATLAB como entorno interactivo y lenguaje de alto nivel para la construcción y ejecución de un script que permita diseñar, entrenar y visualizar distintos modelos de redes neuronales. Esta herramienta admite de forma nativa trabajar con el reconocimiento de patrones, introduciendo técnicas como el aprendizaje por el gradiente del error, generación de curvas ROC y gráficos de confusión, entre otras.
- Debido a la dificultad de conocer de forma anticipada la cantidad de neuronas en la capa oculta que maximiza los resultados obtenidos, fueron contemplados distintos valores arbitrarios para este parámetro. El conjunto de datos utilizado en base a esto es el siguiente:

$$\text{NCP} = \{5; 8; 9; 10; 11; 12; 13; 14; 15; 20; 25; 40; 70; 100; 120\}$$

- La función de activación en la capa oculta tenida en cuenta para este análisis fue la sigmoidea o logística.

El algoritmo utilizado para esta etapa es presentado y explicado en la sección *Apéndice 2: Algoritmo de entrenamiento*.

Por otra parte, ha existido un interés por identificar los rendimientos obtenidos al aplicar distintas restricciones en el conjunto de datos inicial. Para esto, la etapa de entrenamiento fue dividida en dos grandes partes: en una primera instancia se trabajó con los registros de animales que se encontraban dentro de su primer tercio de lactancia (menos de 65 días en lactación, período en que la curva de producción presenta una pendiente ascendente y donde se presuponía la identificación de la presencia del suceso sería factible de realizar más fácilmente), para luego hacerlo con el conjunto completo. A su vez, en relación a la ventana de tiempo utilizada para

clasificar a los datos como positivos, en cada etapa se ha realizado una ejecución independiente del proceso por cada una de las alternativas presentadas a continuación:

- Día en que se registró la presencia de mastitis clínica en el animal
- Día en que se registró la presencia de mastitis clínica en el animal y día previo
- Día en que se registró la presencia de mastitis clínica en el animal y sus dos días previos
- Día previo al que se registró la presencia de mastitis clínica en el animal
- Dos días previos al que se registró la presencia de mastitis clínica en el animal
- Día en que se registró la presencia de mastitis clínica en el animal, sus dos días previos y sus dos días subsiguientes
- Día anterior al día previo en que se registró la presencia de mastitis clínica en el animal

Resulta pertinente aclarar que la ventana de tiempo ha sido ampliada en un máximo de dos días previos desde el momento en el cual se efectuó el diagnóstico debido a que ese período es el mínimo presentado para los distintos agentes desde que la enfermedad es contraída hasta su manifestación como mastitis clínica, fecha indicada en los registros existentes como aparición de la enfermedad. A su vez, el concepto de ventana de tiempo utilizado puede ser comprendido como el período de tiempo que fue analizado para clasificar los datos, teniendo como referencia el momento en que la enfermedad ha sido detectada como mastitis clínica.

Por otra parte, fue expresado como parámetro de comparación entre una y otra red el coeficiente medio de confusión. Este valor se obtiene a partir de la Ecuación 3, donde ric se refiere al número de registros clasificados incorrectamente y rc al total de registros clasificados.

$$c = \frac{ric}{rc}$$

Ecuación 3. Coeficiente de confusión.

Por último, el tipo de entrenamiento utilizado fue de tipo batch, donde todos los valores que componen el conjunto de entrada son pasados a la red antes de que se produzca la actualización en los pesos sinápticos de la misma. El algoritmo de entrenamiento aplicado es el Método del Gradiente Conjugado [16]. Estas

características fueron seleccionadas debido a que presentan generalmente el mejor rendimiento en los problemas de clasificación de patrones, de acuerdo a lo expresado por el fabricante de la herramienta utilizada. A su vez, el entrenamiento también estuvo regido por el método Early Stopping, el cuál detiene este proceso cuando el error de validación aumenta en 6 iteraciones consecutivas.

5. Resultado

De acuerdo a lo expresado anteriormente un total de 14 pruebas fueron realizadas iterando los conjuntos de datos y la configuración utilizada (cantidad de neuronas en la capa oculta, ventana de tiempo, entre otras). En la sección *Apéndice 3: Resultados obtenidos del proceso de entrenamiento* del presente documento, se introducen los resultados de las mismas.

A partir de esto y a fin de obtener un porcentaje de clasificación de animales con mastitis clínica con el menor error posible en el día en el que dicha patología se hace visible, resulta conveniente utilizar la red obtenida en la prueba número 1 (en caso de utilizar datos de animales con menos de 65 días en lactación) o número 8 (en caso de utilizar la totalidad de los registros). En dichos ensayos los resultados de las mismas ofrecen un coeficiente de confusión promedio (*Ecuación 3. Coeficiente de confusión.*) igual a 0,0640 y 0,0687, respectivamente. En términos de sensibilidad (probabilidad de clasificar un evento de interés como mastitis clínica cuando realmente lo es) y especificidad (probabilidad de que cuando el evento de interés no ocurra se lo clasifique como normal o no infectada), los valores arrojados son 98,8% y 88,4%, para la prueba número 1, mientras que para la prueba número 8 fueron 97,7% y 88,6%. Estas variables han sido obtenidas correspondientemente a partir de las ecuaciones presentadas en el *Apéndice 4: Cálculos de especificidad y sensibilidad.*

6. Conclusión

Si bien esto no soluciona la problemática actual, dado que el momento en el que se realiza la detección de la enfermedad se mantendría constante, permite contar con una herramienta de apoyo al personal del tambo del INTA EEA Rafaela de modo que las

pruebas diarias que se realizan en los animales para llevar a cabo dicho diagnóstico sea eliminado o reducido parcialmente a un conjunto de animales más acotado.

En lo que se refiere a los valores de sensibilidad (SE) y especificidad (ES) no existen estándares nacionales o internaciones fijados hasta este momento para la aplicación práctica. No obstante, valores de 70% de SE y 99% de ES son propuestos por Mein y Rasmussen [17]. A su vez, una SE mayor a 80% y una ES cercana al 99% son descritas por la Organización Internacional de Normalización (ISO) en el anexo C de la norma ISO/FDIS 20966 para instalaciones lecheras automatizadas. Contrastando los resultados obtenidos con trabajos semejantes [2, 3], se puede mencionar que los valores de especificidad y sensibilidad han sido levemente superiores en el presente estudio.

Sobre la aplicación de redes neuronales artificiales en esta temática observando los resultados a los que se ha llegado, se puede concluir que son ampliamente utilizables en el contexto trabajado. La implementación de redes Perceptrón Multicapa podría constituir una ayuda para monitorizar la presencia de mastitis en los animales del establecimiento estudiado. Por otra parte, este estudio constituye un aporte en la intersección entre dos áreas del conocimiento, la cual no ha sido ampliamente abordada en la región hasta el día de la fecha.

Bibliografía

- [1] Martin Pol. (2012) <http://www.motivar.com.ar/>. [Online].
<http://www.motivar.com.ar/2013/08/un-tambo-promedio-pierde-35-dias-de-produccion-a-causa-de-mastitis-2/>
- [2] Claudia Kamphuis, *Making sense of sensor data: detecting clinical mastitis in automatic milking systems.*, 2010.
- [3] Zhibin Sun, Sandhya Samarasinghe, and Jenny Jago, "Detection of mastitis and its stage of progression by automatic milking systems using artificial neural networks," 2009.
- [4] Álvaro Castro Ramírez, *Ganadería de leche. Un enfoque empresarial. Producción bovina (tomo 1)*. Costa Rica: Universidad Estatal a Distancia de San José, 2002.
- [5] Richard Bellman, *An Introduction to Artificial Intelligence: Can Computers Think?:* Boyd & Fraser, 1978.

- [6] Carlos N. Corbellini. (2012) Facultad de Agronomía Universidad de Buenos Aires. [Online]. <http://www.agro.uba.ar/sites/default/files/agronomia/la-mastitis-bovina-y-su-impacto-sobre-calidad-de-leche.pdf>
- [7] Simon Haykin, *Neural networks - A comprehensive foundation.*: Prentice Hall: pag. 205-206, 1999, vol. II.
- [8] Roberto Hernández Sampieri, Carlos Fernández Collado, and Pilar Baptista Lucio, *Metodología de la investigación*, Tercera edición ed.: Mc Graw Hill, 2003.
- [9] P. D. P. Wood, *Algebraic model of lactation curve in cattle.*: Nature: pag. 164-165, 1967.
- [10] Daniel Zaborski and Wilhelm Grzesiak, "Detection of difficult calvings in dairy cows using neural classifier," Department of Ruminants Science, West Pomeranian University of Technology in Szczecin, Szczecin, Poland, ISSN: 0003-9438, 2011.
- [11] Javier Chaves, "Calidad de leche y mastitis bovina," 2009.
- [12] Luis F. Calvino, "Mastitis bovina," Rafaela, 2009.
- [13] Charles Alais, *Ciencia de la leche. Principios de técnicas lecheras.*: Reverté S. A. : pag. 464, 1985.
- [14] Alejandra Cuatrin, , 2011.
- [15] Mordecai Avriel, *Nonlinear Programming: Analysis and Methods.*: Dover Publishing, 2003.
- [16] G. A., and M. D. Rasmussen Mein, *erformance evaluation of systems for automated monitoring of udder health: would the real gold standard please stand up? Mastitis Control - From science to practice.*: Wageningen Academic Publishers, 2008.
- [17] S. & M. Gallardo Valtorta, "El estrés por calor en producción lechera," 1996.
- [18] G.L. Hahn, "Environmental management for improved livestock performance, health and well-being," 1995.
- [19] E. C. Thom, "The discomfort index," 1959.
- [20] H.D. Johnson, H.H. Kibler, A.C. Ragsdale, and I.L. & Shanklin, M.D. Berry, "Role of heat tolerance and production level in responses of lactating Holsteins to various temperature-humidity conditions.," 1961.
- [21] S. Samarasinghe and J.G.H. H, J.G.H. Hickford, and M.G. López-Benavides, *The use of artificial neural networks to diagnose mastitis in dairy cattle.* Lincoln University, Canterbury, 2003.

Apéndice 1: Cálculo de estrés calórico

La temperatura ambiente adecuada de la vaca lechera se encuentra aproximadamente entre los 5 y 21 °C, siempre que no existan otros factores limitantes que pudiesen afectar este umbral [18]. A su vez, condiciones ambientales con temperaturas elevadas pueden afectar de forma adversa a la salud y bienestar del animal [19].

Esta variable será representada mediante lo que se conoce como el Índice de Temperatura y Humedad (ITH) desarrollado por Thom [20]. Este valor se obtiene a partir de la Ecuación 4, donde T_a representa la temperatura del aire (°C) y HR la humedad (%).

$$ITH = (1,8 T_a + 32) - \left(0,55 - 0,55 \frac{HR}{100}\right) (1,8 T_a - 26)$$

Ecuación 4. Índice de temperatura y humedad.

Del resultado alcanzado en el cálculo de este índice será interpretada como situación de estrés aquellos casos donde se supere el valor crítico de 72, determinado por Jonhson et al. [21].

Apéndice 2: Algoritmo de entrenamiento

A continuación se introduce el script generado para realizar el entrenamiento con la herramienta MATLAB.

```
% Create a Pattern Recognition Network
maxPerformance = 500000;

for index3=1:3

    %Se utilizan distintos ratios para crear diversos conjuntos
    if index3==1
        divideRatio = 80;
        divideRatioRemaining = 10;
    elseif index3==2
        divideRatio = 70;
        divideRatioRemaining = 15;
    else
        divideRatio = 60;
        divideRatioRemaining = 20;
    end

    %Se utilizan distintas cantidades de neuronas en la capa oculta
    for index15=1:15
        if index15==1
            hiddenLayerSize = 5;
        elseif index15==2
            hiddenLayerSize = 8;
        elseif index15==3
            hiddenLayerSize = 9;
        elseif index15==4
            hiddenLayerSize = 10;
        elseif index15==5
            hiddenLayerSize = 11;
        elseif index15==6
            hiddenLayerSize = 12;
        elseif index15==7
            hiddenLayerSize = 13;
        elseif index15==8
            hiddenLayerSize = 14;
        elseif index15==9
            hiddenLayerSize = 15;
```

```

elseif index15==10
    hiddenLayerSize = 20;
elseif index15==11
    hiddenLayerSize = 25;
elseif index15==12
    hiddenLayerSize = 40;
elseif index15==13
    hiddenLayerSize = 70;
elseif index15==14
    hiddenLayerSize = 100;
elseif index15==15
    hiddenLayerSize = 120;
end

clear net

%Se obtienen datos negativos de forma aleatoria.
negativeData = datasample(negValues,length(posValues(:,1)));

allData = cat(2, negativeData',posValues');
allData =
datasample(allData',length(posValues(:,1)) * 2);

inputs = allData([1,2,3,4,5,6,7,8,9,10,11,12],:);
targets = allData(13,:);

net = patternnet(hiddenLayerSize);

% Choose Input and Output Pre/Post-Processing Functions
net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};

% Setup Division of Data for Training, Validation, Testing
net.divideFcn = 'dividerand'; % Divide data randomly
net.divideMode = 'sample'; % Divide up every sample
net.divideParam.trainRatio = divideRatio/100;
net.divideParam.valRatio = divideRatioRemaining/100;
net.divideParam.testRatio = divideRatioRemaining/100;

net.layers{2}.transferFcn = 'tansig';
net.trainFcn = 'trainscg'; % Scaled conjugate gradient

% Choose a Performance Function
net.performFcn = 'mse'; % Mean squared error

% Choose Plot Functions
net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...

```

```

    'plotregression', 'plotfit'});

% Train the Network
net.trainParam.epochs = 300;

for index5=1:5
    %Re-entrenar la red y evaluar nuevamente
    [net,tr] = train(net,inputs,targets);

    %Evaluar la red
    outputs = net(inputs);
    errors = gsubtract(targets,outputs);
    performance = perform(net,targets,outputs);

    cNetSum = 0;
    sensibility = 0;
    specificity = 0;

    %IterationsNumber depende de la relación entre los valores positivos y negativos
    iterationsNumber = round(length(negValues(:,1))/length(posValues(:,1)));

    for indexIN=1:iterationsNumber
        % Se obtienen datos negativos de forma aleatoria.
        negativeData = datasample(negValues,length(posValues(:,1)));

        allData = cat(2, negativeData',posValues');
        allData = datasample(allData',length(posValues(:,1)) * 2)';

        inputs = allData([1,2,3,4,5,6,7,8,9,10,11,12],:);
        targets = allData(13,:);

        netOutputs = sim(net, inputs);

        [c,cm,nd,per] = confusion(targets,netOutputs);
        cNetSum = cNetSum + c;

        sensibility =
sensibility + (cm(2,2) / (cm(2,2) + cm(2,1)));
        specificity =
specificity + (cm(1,1) / (cm(1,1) + cm(1,2)));
    end

    %Guardar mejor performance obtenida
    if (maxPerformance > (cNetSum/iterationsNumber))
        maxPerformance = cNetSum/iterationsNumber;
        maxDivideRatio = divideRatio;
        maxHiddenLayerSize = hiddenLayerSize;
    end

```

```

        maxNet = net;
        maxSensitivity = sensibility/iterationsNumber * 100;
        maxSpecificity = specificity/iterationsNumber * 100;
    end
end
end
end

% View the Network
view(maxNet)

```

Dicho código puede ser explicado de la siguiente forma:

En un primer momento se disponen distintos ciclos que permiten ir alternando las diferentes características de la red a entrenar y la forma en la que esa actividad es llevada a cabo (cantidad de neuronas en la oculta y tamaños de los distintos conjuntos de datos). Luego, se conforma el set de datos que se utiliza tomando la totalidad de los registros que indican la presencia de la enfermedad (valores positivos) y una cantidad igual a esta de valores negativos (tomados de forma aleatoria sobre los datos de este tipo). Existe un ciclo de repeticiones especificado que permite en distintas iteraciones tomar todos los datos disponibles, repitiéndose en cada caso los valores positivos e intercambiando los negativos.

Por cada configuración lograda a lo largo del código (cantidad de neuronas en la capa oculta, así como también los porcentajes de los conjuntos de datos) se realiza un entrenamiento para una nueva red, el cual se repite un número de veces a fin de aproximar los pesos sinápticos correctos. A su vez, por cada selección de datos conformada se simulan los resultados obtenidos para esa red en particular. Dichos resultados se comparan red a red, obteniendo finalmente aquella que obtuvo los mejores valores en cuanto al porcentaje de clasificación incorrecta.

Apéndice 3: Resultados obtenidos del proceso de entrenamiento

La Tabla 1 exhibe para cada uno de los distintos ensayos los parámetros utilizados (*) y los resultados obtenidos (**).

Nro. prueba (*)	Datos de entrada (filtros por días en lactación) (*)	Ventana de tiempo (*)	Nro. neuronas en la capa oculta (*)	% conjuntos de estimación/validación/prueba (*)	Coef. de confusión promedio (**)
1	Primer tercio	I	25	80/10/10	0,0640
2	Primer tercio	II	12	80/10/10	0,1717
3	Primer tercio	III	70	80/10/10	0,1897
4	Primer tercio	IV	13	80/10/10	0,2315
5	Primer tercio	V	40	70/15/15	0,2146
6	Primer tercio	VI	40	60/20/20	0,2144
7	Primer tercio	VII	12	80/20/20	0,2560
8	Completo	I	25	60/20/20	0,0687
9	Completo	II	9	80/20/20	0,2306

10	Completo	III	12	70/15/15	0,2581
11	Completo	IV	11	70/15/15	0,2554
12	Completo	V	25	80/10/10	0,2906
13	Completo	VI	15	70/10/10	0,2649
14	Completo	VII	15	70/10/10	0,3293

Tabla 1. Resultados de la fase de entrenamiento.

De acuerdo a lo definido anteriormente, se puede observar la forma en que se han alternado los parámetros utilizados para las distintas pruebas, variando el conjunto de datos y la longitud de la ventana de tiempo utilizada. En relación a esto, se observan en las columnas restantes las configuraciones de las redes que mejores resultados entregaron para cada alternativa.

Apéndice 4: Cálculos de especificidad y sensibilidad

A través de la Ecuación 5. Cálculo de sensibilidad. y la Ecuación 6. Cálculo de especificidad. se obtienen los valores de especificidad y sensibilidad, los cuáles han sido fundamentales para otorgar una magnitud a los resultados de las distintas pruebas.

$$sensibilidad = \frac{verdaderos\ positivos}{verdaderos\ positivos + falsos\ negativos}$$

Ecuación 5. Cálculo de sensibilidad.

$$especificidad = \frac{verdaderos\ negativos}{verdaderos\ negativos + falsos\ positivos}$$

Ecuación 6. Cálculo de especificidad.