

**INSTITUTO TECNOLÓGICO DE COSTA RICA
CAMPUS TECNOLÓGICO LOCAL SAN CARLOS**

**DETECCIÓN DE MASTITIS SUBCLÍNICA EN VACAS LECHERAS
USANDO MODELOS ALOMÉTRICOS Y ALGORITMOS DE
INTELIGENCIA ARTIFICIAL, SAN CARLOS, COSTA RICA**

Trabajo Final de Graduación presentado a la Escuela de Agronomía
como requisito parcial para optar al grado de Licenciatura
en Ingeniería en Agronomía

VERNY VICENTE ESTRADA CARVAJAL



2019

Carrera de Ingeniería en Agronomía
Tecnológico de Costa Rica
Sede Regional San Carlos

**DETECCIÓN DE MASTITIS SUBCLÍNICA EN VACAS LECHERAS
USANDO MODELOS ALOMÉTRICOS Y ALGORITMOS DE
INTELIGENCIA ARTIFICIAL, SAN CARLOS, COSTA RICA**

VERNY VICENTE ESTRADA CARVAJAL

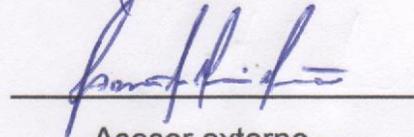
Aprobado por los miembros del Tribunal Evaluador:

Ing. Agr. Wilfrido Paniagua Madrigal, MGA.



Asesor principal

M. V. Rafael Alberto Molina Montero, MBA.



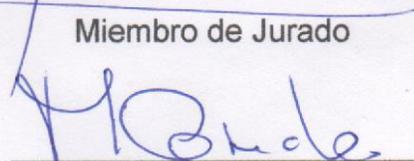
Asesor externo

M. V. Jaime Galindo Badilla, M.Sc.



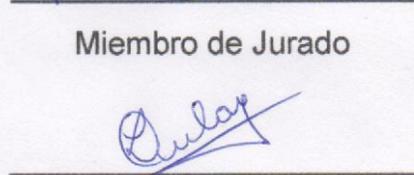
Miembro de Jurado

Ing. Fo. Marlen Camacho Calvo, M.Sc.



Miembro de Jurado

Ing. Agr. Zulay Castro Jiménez, MGA.



Coordinadora Trabajos
Finales de Graduación

Ing. Agr. Sergio Torres Portuguez, M. Sc.



Director de la
Escuela de Agronomía

DEDICATORIA

A Dios, por guiar mis pasos para la conclusión de este documento y mi carrera de Licenciatura. También dedico esto a mis padres, Eddy Estrada Jara y Marianela Carvajal Blanco, y a mi novia, Margarita Gamboa, quienes me apoyaron, motivaron e impulsaron en mi carrera.

Con mucho cariño.

Verny Estrada Carvajal

AGRADECIMIENTOS

Primeramente, agradezco a Dios por haber guiado mis pasos para concluir este proyecto y mi carrera con éxito.

Agradezco a mi familia por el gran apoyo que me brindó durante los años de estudio, a mis padres por ser un gran ejemplo para seguir.

Mi agradecimiento también para mi asesor principal, Ing. Wilfrido Paniagua Madrigal; MGA y para la profesora Ing. Marlen Camacho Calvo; MSc., por el aporte de conocimiento científico y técnico, y por el compromiso adquirido durante todo el proceso de la tesis.

Agradecimiento especial a los propietarios y el personal de finca Ganadera La Uno S.A, por su colaboración con el proyecto y por brindarme acceso a sus instalaciones y equipo para realizar la investigación.

También al M.V. Rafael Alberto Molina Montero; MBA. y a la Cooperativa de Productores de Leche (COOPROLE) por los aportes en la asesoría técnica, así como materiales y equipos necesarios para el procesamiento de muestras.

Un agradecimiento muy especial a los amigos y compañeros que colaboraron de una u otra forma para el enriquecimiento del documento. Entre ellos, le agradezco a el señor Nelson Gamboa Guzmán por su aporte como especialista de estilo, a Augusto César Gamboa Avellán por aplicar sus conocimientos de la carrera de Ingeniería en Sistemas Informáticos de la Universidad Nacional, a Andrey Vega Alfaro, como colega de la carrera de Ingeniería en Agronomía del Tecnológico de Costa Rica y estudiante de maestría en la Universidad de Wisconsin, Madison. También a Margarita Gamboa Avellán por ser mi gran motivación en la conclusión de este trabajo.

Finalmente, a todos (profesores, compañeros, amigos, familiares) los que de alguna manera me ayudaron para poder culminar mi carrera.

Atentamente:

Verny Estrada Carvajal

TABLA DE CONTENIDO

DEDICATORIA.....	i
AGRADECIMIENTOS	ii
LISTA DE CUADROS.....	vi
LISTA DE FIGURAS.....	vii
RESUMEN	ix
ABSTRACT	x
1 INTRODUCCIÓN.....	1
1.1 Justificación	1
1.2 Objetivo general	3
1.3 Objetivos específicos	3
1.4 Hipótesis de investigación.....	3
2 REVISIÓN DE LITERATURA	4
2.1 Generalidades de la mastitis	4
2.1.1 Agente causal.....	4
2.1.2 Patogénesis	5
2.1.3 Epidemiología y etiología	6
2.1.4 Síntomas.....	7
2.1.5 Cambios en la composición de la leche.....	7
2.2 Impacto económico de la mastitis.....	7
2.3 Métodos de detección de mastitis	9
2.3.1 Conteo de células somáticas (CCS)	10
2.3.2 Conductividad eléctrica (CE)	10
2.3.3 Sistemas en línea.....	12
2.3.4 Otros métodos.....	12
2.4 Modelos alométricos para la detección de mastitis.....	12

2.5	Inteligencia artificial en la detección de mastitis	13
2.6	Comparación de métodos de diagnóstico.....	15
2.6.1	Sensibilidad.....	15
2.6.2	Especificidad.....	16
3	MATERIALES Y MÉTODOS.....	17
3.1	Descripción del sitio de estudio	17
3.1.1	Protocolo de ordeño	19
3.1.2	Protocolo de lavado del equipo de ordeño.....	21
3.2	Periodo de estudio	21
3.3	Material experimental	21
3.4	Unidad muestral	22
3.5	Variables regresoras y de respuesta	22
3.6	Categorías de estratificación del muestreo.....	23
3.7	Metodología de toma de datos	23
3.8	Análisis de datos	26
3.8.1	Análisis exploratorio	26
3.8.2	Modelos alométricos (MR).....	26
3.8.3	Algoritmos de aprendizaje automático (<i>Machine learning</i>).....	27
3.8.4	Sensibilidad y especificidad.....	28
4	RESULTADOS Y DISCUSIÓN	31
4.1	Descripción de la población.....	31
4.1.1	Impactos económicos sobre la explotación	32
4.2	Relación entre las variables evaluadas	34
4.3	Comparación entre las variables de estudio	35
4.4	Selección del modelo	40

4.4.1	Modelos alométricos.....	40
4.4.2	Algoritmos de <i>Machine Learning</i>	41
4.4.3	Prueba de sensibilidad (SEN) y especificidad (ESP)	42
5	CONCLUSIONES	48
6	RECOMENDACIONES.....	49
7	BIBLIOGRAFÍA.....	50
8	ANEXOS	58

LISTA DE CUADROS

Cuadro	Título	Página
1	Interpretación de los valores de conductividad eléctrica absoluta (CEA) y la conductividad eléctrica diferencial (CED) en la leche.	11
2	VARIABLES regresoras y de respuesta evaluados en el estudio de detección de mastitis subclínica, San Carlos, 2017.	22
3	Comportamiento general del conteo celular somáticas (CCS) según muestreos semanales, durante el estudio. Finca Ganadera La Uno S.A, San Carlos, 2017.	31
4	Comparación de la media y desviación estándar de la CE, PL y CCS entre vacas sanas (0) y enfermas (1) ($p < 0.05$). Finca La Uno S.A, San Carlos, 2017.	36
5	Comparación de criterios de selección de modelos de regresión (MR) y sus respectivas ecuaciones, obtenidas durante el estudio. Finca Ganadera La Uno S.A, San Carlos, 2017.	40
6	Análisis de sensibilidad (SEN), especificidad (ESP), precisión y tasa de error, para todos los modelos alométricos y de <i>machine learning</i> . Finca Ganadera La Uno S.A, San Carlos, 2017.	43

LISTA DE FIGURAS

Figura	Título	Página
1	Infraestructura para la alimentación de las vacas en producción. Finca Ganadera La Uno S.A (2017).	17
2	Vista general de la sala de ordeño en finca Ganadera La Uno S.A, San Carlos, 2017.	18
3	Pantalla de control de ordeño, producción de leche (Kg), conductividad eléctrica (mS/cm) y diferencia (%) respecto a ordeños anteriores. Finca Ganadera La Uno S.A, San Carlos, 2017.	19
4	Protocolo de ordeño, a) despunte de la vaca: extraer primeros chorros, b) máquinas retiradas automáticamente y pezón recubierto con sellador. Finca Ganadera La Uno S.A, San Carlos, 2017.	20
5	a) Consola y retirado automático - TOUCHPOINT™ para control de ordeño, visualización e ingreso de información individual; b) vista general del equipo automático de ordeño; c) máquina de ordeño (1), frasco recolector (2), flujómetro – SMARTCONTROL™ METER - US (3), medidor de CE - SMARTFLO™ (4) y cable de retirado automático (5). Finca Ganadera La Uno S.A, San Carlos, 2017.	24
6	a) Identificación y colocación de la bolsa para muestreos de leche en el frasco recolector de la muestra, b) filtrado y vaciado de la leche en la bolsa de muestreo. Finca Ganadera La Uno S.A, San Carlos, 2017.	25
7	Relación entre el resultado del modelo y análisis de laboratorio para la obtención de parámetros de comparación (SEN, ESP, precisión y tasa de error).	29
8	Comportamiento del ingreso bruto (colones) por vaca por día según score de células somáticas (SCCS) del hato en estudio. Finca Ganadera La Uno S.A, San Carlos, 2017.	33
9	Análisis de componentes principales (ACP) para las variables regresoras y dependientes. a) considera la variable CE, b) excluye la variable CE. Finca Ganadera la Uno S.A, San Carlos, 2017.	35

- 10** Comportamiento de tres variables continuas (score de células somáticas: SCCS, conductividad eléctrica: CE, y producción de leche: PL) según las variables categóricas (número de partos: NPc, días en lactación: DELc y Edad: EDc). a) SCCS vrs NPc, b) SCCS vrs DELc, c) SCCS vrs EDc, d) CE vrs NPc, e) CE vrs DELc, f) CE vrs EDc, g) PL vrs NPc, h) PL vrs DELc, i) PL vrs EDc. Finca Ganadera La Uno S.A, San Carlos, 2017. **37**
- 11** Curvas de sensibilidad (SEN) y especificidad (ESP) del modelo MLRG. Se incluye el *threshold* y se indica la ESP a un 80% de ESP fija. Línea de corte vertical indica el valor de 4 en el eje “x”. Finca Ganadera La Uno S.A, San Carlos, 2017. **44**
- 12** Curva ROC para el modelo MLRG, se incluye el área, error y p-valor de la curva. Se construye con 1-ESP y la SEN, expresados en porcentaje (%). Finca Ganadera La Uno S.A, San Carlos, 2017. **46**

RESUMEN

El estudio sobre la detección de mastitis subclínica se realizó en la Marina de San Carlos en finca Ganadera La Uno S.A (coordenadas CRTM05 458742.53 E y 1147645.93 N). El objetivo fue aplicar y comparar métodos para la detección de mastitis subclínica, con base en modelos alométrico y algoritmos de inteligencia artificial. Los datos fueron recolectados con un sistema de ordeño automatizado con mediciones en línea. Las vacas enfermas fueron definidas como aquellas que presentaron conteo de células somáticas (CCS) mayor a 200 000 cs/ml o un score de células somáticas (SCCS) de 4 o mayor. La mejor variable de respuesta fue el score de células somáticas, por su distribución normal y buen ajuste en los modelos. Como variables predictoras de SCCS se utilizaron la conductividad eléctrica (CE), la producción de leche (PL), e indicadores técnicos de las vacas como número de partos (NP), días en lactación (DEL), edad (ED) y estado reproductivo (ER). Los modelos alométricos fueron lineales o linealizados, y mostraron valores de R^2 por debajo de 0,32. Además, el valor de predicción de estos fue deficiente en términos de sensibilidad (SEN) y especificidad (ESP), ya que la especificidad fue de 55% cuando la sensibilidad se fijó en 80% y, el área bajo la curva ROC de estos modelos se aproximó a 75%. Los resultados mostraron que el algoritmo de clasificación de Naive bayes no fue un buen estimador de mastitis subclínica, con valores de SEN y ESP similares a los modelos alométricos. Por último, se concluyó que el mejor predictor fue el modelo de regresión lineal que aplicó *machine learning*, con ESP (a 80% de SEN) y áreas bajo la curva ROC de 84,3% y 92,1%, respectivamente. Se demostró que existe un gran potencial de los algoritmos de inteligencia artificial para la detección de mastitis subclínica en hatos lecheros.

ABSTRACT

A mastitis detection trial was performed at Ganadera La Uno S.A, in la Marina de San Carlos (CRTM05 coordinates 458742.53 E y 1147645.93 N). The aim of this investigation was to develop and compare mastitis detection models based on allometric models and artificial intelligence algorithms. Data was collected from an automated milking system equipped with online measurements. Diseased cows were defined either as those which recorded somatic cell count (SCC) greater than 200 000 cs/ml or as those with a somatic cell count score (SCCS) of four or greater. The best response variable in was SCCS due to its compliance with the normal distribution and adequate fit in the models. Electric conductivity (CE), milk yield (PL), days in milk (DEL), age (ED), reproductive condition (ER) and lactation number (NP) were used as predictive variables of SCCS. Allometric models were linear or linearized and showed R² values lower than 0.32. Moreover, predictive value of these models was inadequate in terms of sensibility (SEN) and specificity (ESP), since the specificity was 55% when the sensitivity was fixed at 80% and the area under the ROC curve of these models approached 75%. Results do not support the use of a Naïve Bayes classification algorithm to predict mastitis because sensitivity and specificity values were inadequate. The best predictive model was a linear regression model combined with machine learning, with values of ESP and ROC area curve of 84.3 and 92.1% respectively (80% fixed sensibility). This work presents evidence of the potential value of artificial intelligence models in mastitis detection in dairy farms.

1 INTRODUCCIÓN

1.1 Justificación

La mastitis es una reacción inflamatoria del tejido mamario, contra el daño producido por toxinas e irritantes liberados por bacterias que colonizan la ubre. Esta inflamación lleva a un incremento en la producción de glóbulos blancos (células somáticas) que erradican los irritantes, reparan el tejido dañado y recuperan la ubre infectada. En adición, la calidad de la leche se ve afectada negativamente por conteos altos de células somáticas (Viguiet *et al.* 2009).

La mastitis induce una reducción en la producción de leche diaria, que alcanza hasta un 75% del total de leche producido por vaca (Guimarães *et al.* 2017). Esta enfermedad se cataloga como uno de los principales rubros de pérdidas en fincas lecheras (Nielsen 2009, Guimarães *et al.* 2017) y compromete seriamente el bienestar animal. La mastitis, además, se caracteriza por niveles de incidencia alta, inclusive en las granjas con rigurosos programas de control de infección por mastitis (Fernández *et al.* 2008, Mora *et al.* 2015). Es también la enfermedad más influyente sobre los costos de producción de leche bovina (Batavani *et al.* 2007, Sharma *et al.* 2011, Zhao y Lacasse 2014, Gráf y Mikó 2015).

La inocuidad de la leche y el bienestar animal en fincas lecheras pueden mejorar a través de sistemas de detección temprana de enfermedades, como la mastitis (De Mol y Ouweltjes 2001). Los métodos más comunes son la prueba de California para mastitis (CMT, por sus siglas en inglés), el conteo de células somáticas (CCS) y el cultivo de bacterias (Sharma *et al.* 2010). Sin embargo, estos métodos son costosos y demandantes de tiempo, por lo cual son muchas veces evitados en granjas lecheras.

La incorporación de sistemas de ordeño automatizados reduce drásticamente la interacción del operario con el animal, lo que hace necesario la incorporación de sistemas de detección en línea, utilizando parámetros de fácil medición *in situ* para detectar mastitis (Momena *et al.* 2017). Estos métodos se basan principalmente en la medición de la conductividad eléctrica (CE) de la leche (Elizalde *et al.* 2009, Kaşıkçı *et al.* 2012, Ribeiro *et al.* 2016), la producción de leche y otras variables propias de cada animal. Un

conteo alto en CCS coincide con una mayor concentración de iones (Na^+ , K^+ y Cl^-) en disolución (Ribeiro *et al.* 2016). Esta variable (CE) ha sido utilizada recientemente en múltiples estudios como una forma rápida para la detección de mastitis y sistemas de ordeño automatizados (Ribeiro *et al.* 2016; Momena *et al.* 2017).

Por su versatilidad, el conteo de células somáticas (CCS) ha sido utilizado en muchos estudios sobre detección de mastitis (Polat *et al.* 2010, Sharma *et al.* 2010, Ribeiro *et al.* 2016). Se utiliza como una medición fiable que se aproxima al estado de salud de la ubre (Zambrano y de Pinho, 2008). De esta manera, es una variable respuesta útil para aproximarse a una infección de mastitis, además permite comparar, depurar y validar herramientas de predicción que utilizan variables predictoras como la CE, producción de leche y otros factores de corrección propios de cada animal. Con lo cual se pueden obtener métodos precisos, que disminuyan los costos en diagnóstico de la enfermedad y que, potencialmente, previene pérdidas en los rendimientos por la detección temprana de la enfermedad.

A pesar de que varios estudios han intentado relacionar la mastitis con CE por medio de correlaciones simples, regresiones locales y aleatorias (Norberg, 2004; Cavero, 2007) y otros métodos para el desarrollo de modelos predictivos, la utilidad de estos modelos es cuestionable por su baja sensibilidad y especificidad.

En contraste, los métodos de Redes Neuronales Artificiales, son excelentes candidatos para el desarrollo de herramientas predictivas de mastitis, por tres razones: i) son capaces de utilizar amplios y complejos bagajes de datos ii) no dependen de los supuestos estadísticos paramétricos o distribuciones iii) son métodos precisos que ayudan a la toma rápida de decisiones en finca, que tienen repercusiones técnicas o económicas sobre el ingreso del productor de leche (Shahinfar, *et al.* 2014, Shinea *et al.* 2018) La implementación de modernos equipos de ordeño, junto a un adecuado análisis de la información, puede ser clave para el crecimiento y actualización del sector lechero en Centroamérica. El objetivo de este proyecto fue desarrollar modelos alométricos y algoritmos de inteligencia artificial para la detección de mastitis subclínica, con base en la conductividad eléctrica, la producción de la leche e indicadores técnicos de la vaca.

1.2 Objetivo general

Desarrollar modelos alométricos y algoritmos de inteligencia artificial para la detección de mastitis subclínica, utilizando variables de rápida medición en la línea de leche e indicadores técnicos de la vaca.

1.3 Objetivos específicos

1. Describir el comportamiento de la salud de la glándula mamaria y niveles de producción de la población evaluada utilizando los datos del muestreo y sus implicaciones económicas.
2. Relacionar el score de células somáticas con la conductividad eléctrica, producción y otros indicadores técnicos.
3. Comparar el score de células somáticas, la producción de leche y la conductividad eléctrica con otros indicadores técnicos
4. Validar modelos alométricos de predicción de score de células somáticas en función de la conductividad eléctrica, la producción de leche y otros indicadores técnicos.
5. Entrenar algoritmos de inteligencia artificial para la predicción mastitis subclínica, en función de la conductividad eléctrica, la producción láctea y otros indicadores técnicos.
6. Comparar mediante un análisis de sensibilidad y especificidad, los modelos alométricos y los algoritmos de inteligencia artificial.

1.4 Hipótesis de investigación

Los modelos o herramientas de predicción desarrollados a partir de la conductividad eléctrica, la producción láctea e indicadores técnicos, son adecuados para predecir mastitis subclínica.

2 REVISIÓN DE LITERATURA

2.1 Generalidades de la mastitis

La inflamación de la glándula mamaria en vacas lecheras es conocida como mastitis, y puede ser ocasionada por agentes patógenos, traumas o toxinas (Sharma *et al.* 2011, Zhao y Lacasse 2014, Mora *et al.* 2015, Alfonso *et al.* 2017). Se conocen hasta 137 especies de bacterias que atacan las vacas lecheras (Alfonso *et al.* 2017), afectando negativamente el bienestar de los animales y las ganancias de la industria láctea mundial (Yang *et al.* 2016, Momena *et al.* 2017). Por éstas y otras razones, la enfermedad es considerada la más importante de la producción lechera (Viguier *et al.* 2009, Yang *et al.* 2016).

De acuerdo con Viguier *et al.* (2009), la severidad de la mastitis puede ser clasificada en mastitis subclínica (MSC), mastitis clínica (MC) y mastitis crónica. Esta severidad puede variar en cada animal según sea el agente causal, edad del animal, etapa en la lactación de la vaca y su estado inmunológico (Mora *et al.* 2015). La MSC se ha caracterizado por su difícil detección, debido a que no presenta indicadores visibles, y es la de mayor impacto en los costos. La mastitis crónica es la forma menos frecuente de mastitis, no obstante, resulta en la inflamación persistente de la glándula mamaria (Viguier *et al.* 2009).

La mastitis subclínica se ha relacionado con cambios en la consistencia de la leche, así como en la reducción de la producción, afectando el rendimiento industrial y propiciando una posible contaminación por patógenos (Sharma *et al.* 2011, Digiovani *et al.* 2016). Por otro lado, la mastitis clínica presenta síntomas como enrojecimientos de la ubre, inflamación y aumento de la temperatura del cuarto afectado y cambios en las características de la leche (Rees *et al.* 2017).

2.1.1 Agente causal

La mastitis se puede desencadenar por muchas causas, la principal es una amplia gama de cepas bacterianas, aunque también puede ser producida por virus, algas, hongos (Viguier *et al.* 2009) y micoplasmas (*Mycoplasma bovis*) (Nicholas *et al.* 2016) con menor frecuencia. Lee *et al.* (2008) desarrolló un biochip para identificar los

principales agentes causales de la mastitis en muestras de leche de manera más rápida. Entre los agentes más importantes se mencionan: *Corynebacterium bovis*, *Mycoplasma bovis*, *Staphylococcus aureus*, *Streptococcus* spp; *S. agalactiae*, *S. bovis*, *S. dysgalactiae*, y *S. uberis* (Lee *et al.* 2008, Koskinen *et al.* 2009) causantes de más del 95% de todas las infecciones subclínicas y clínicas (Koskinen *et al.* 2009)

Además, se ha demostrado que *Staphylococcus aureus* es la bacteria que infecta la ubre con mayor prevalencia (Alfonso *et al.* 2017; Guimarães *et al.* 2017) y es de los motivos más importantes de sacrificio de vacas productoras. Incluso se ha incursionado en la infección experimental de vacas secas de la raza Holstein con *S. aureus*, en busca de terapias como el uso de lisostatina recombinante, fusionada con un dominio de transducción de proteínas (rLYS-PTD) (Hoernig *et al.* 2016).

2.1.2 Patogénesis

La patogénesis de la mastitis ha sido descrita en detalle por Viguier *et al.* (2009): La glándula mamaria cuenta con una serie de barreras físicas y químicas que protegen ante la entrada de patógenos. El esfínter, recubierto por queratina cerosa del epitelio escamosos estratificado, cierra el canal del pezón. Además, secreta ácidos grasos de cadena larga que funcionan como antimicrobianos. Este componente, aunque forma parte importante de la barrera de protección, tienen un efecto limitado. Una vez dentro de la ubre, el agente causal de la mastitis debe evadir los mecanismos de defensa celular y humoral de la vaca, su multiplicación comienza por toda la glándula. Las bacterias liberan toxinas que inducen a las células epiteliales a liberar quimioattractantes como citoquinas, factor de necrosis humoral- α (TNF α), interleucina (IL) -8, IL-1, eicosanoides como prostaglandina F $_{2\alpha}$ (PGF $_{2\alpha}$), radicales de oxígeno, y proteínas de fase aguda (APPs), como haptoglobina (Hp) y amiloide sérico a (SAA). Esta situación hace que las células de defensa circulantes se vean atraídas, en especial los neutrófilos polimorfonucleares (PMN), al punto de infección. La forma de controlar las bacterias por parte de los PMN es por medio de su destrucción y/o encapsulamiento en medios aeróbicos o anaeróbicos. Los PMN tienen gránulos intracelulares que contienen péptidos bactericidas, proteínas, enzimas (mieloperoxidasa) y proteasas neutras y ácidas (elastasa, catepsina G, B y D). Estos compuestos destruyen las bacterias y

algunas de las células epiteliales, de manera que se da una reducción de la producción de leche y liberación de enzimas como la N-acetil- β -D-glucosaminidasa (NAGasa) y lactato deshidrogenasa (LDH). La apoptosis es el método por el cual se destruyen la mayoría de PMN, una vez cumplida su función. Los macrófagos terminan de envolver e ingerir los PMSs restantes. Finalmente, las células epiteliales mamarias muertas y leucocitos muertos son agregados en la leche para ser expulsados, esto resulta en el aumento del CCS.

2.1.3 Epidemiología y etiología

La mastitis, al igual que muchas enfermedades de los animales, es considerada multifactorial, ya que requieren que se complemente una serie de elementos para desarrollarse. Estos elementos componen el “triángulo de la enfermedad”, que está determinado por: el hospedero (ubre de la vaca), ambiente (manejo y clima) y patógeno (agente causal). Cuando estos elementos tienen condiciones específicas que hacen susceptible al animal y el patógeno tiene medios para ingresar por el canal del pezón, es cuando se inicia de la enfermedad (Ferrero 2014).

Mora *et al.* (2015) menciona que existen factores fijos con efecto significativo para casos de incidencia de mastitis, dentro de los cuales menciona la zona agroecológica, el grupo racial, el año, número y mes de parto, periodo de lactancia, duración y producción de leche en lactancia previa e historial previo de mastitis. Por otro lado, Pedraza (1991) menciona que la estación climática es importante para detectar diferencias en el rendimiento entre las vacas sanas y las enfermas; en la época seca, la brecha entre éstas disminuye, debido a que las sanas descienden a su rendimiento más bajo y las enfermas alcanzan su mayor productividad por las condiciones favorables propias de la época donde los cuartos enfermos se afectan en menor grado.

Los factores ambientales son un pilar importante en el desarrollo de las enfermedades. Para los animales en confinamiento se ha demostrado que el uso de una cama adecuada puede disminuir el riesgo de infección por mastitis. Rowbotham y Ruegg (2016), han demostrado que el uso de camas de arena nueva de lecho profundo son las más efectivas para que las vacas primíparas puedan resistir más tiempo la incidencia de mastitis.

2.1.4 Síntomas

Como ya se ha mencionado, la sintomatología de la mastitis depende del tipo. Cuando se trata de una MSC, no es posible identificarla mediante síntomas visibles (Viguiet *et al.* 2009). No obstante, cuando los casos son de MC, se caracteriza por una respuesta inflamatoria de la glándula mamaria (Oviedo-Boyso *et al.* 2007, Rees *et al.* 2016, Yang *et al.* 2016), causada por cambios metabólicos y fisiológicos, traumatismos o, más frecuentemente, microorganismos patógenos contagiosos o ambientales. En la respuesta inflamatoria, otros síntomas detectables son los cambios en las características de la leche, hinchazón, aumento de la temperatura del cuarto afectado (Rees *et al.* 2017) y disminución de la producción de leche (Lee *et al.* 2008, Nielsen 2009, Guimarães *et al.* 2017).

2.1.5 Cambios en la composición de la leche

Se ha demostrado que la mastitis afecta la proporción de los componentes de la leche, entre ellos la proteína, lactosa y grasa, afectando la calidad y rendimiento industrial (Vásquez *et al.* 2014). Estos cambios en la composición de la leche pueden ser utilizados para medir indirectamente el estado de afectación de la glándula mamaria. Entre algunas mediciones que se han realizado, se reportan: la conductividad eléctrica, el conteo de células somáticas, el contenido de cloruro y lactosa para detectar cuadros de MC o MSC (Ribeiro *et al.* 2016).

Se han reportado cambios en la composición de las proteínas del plasma y de la leche, como respuesta a la infección de la glándula mamaria por *Escherichia coli* en vacas Holstein francés. Específicamente, mencionan aumentos en la albúmina y caseínas variantes en casos leves de mastitis, mientras que en casos severos de infección se observa la disminución de la caseína, acompañada de aumentos en antimicrobianos y proteínas de fase aguda. También se ha asociado la alfa-1-ácido glicoproteína y catelicidinas con mastitis severa (Yang *et al.* 2016).

2.2 Impacto económico de la mastitis

La infección de la glándula mamaria es considerada una de las enfermedades más importantes de la industria lechera, caracterizada por valores altos de incidencia, incluidas las granjas con programas de control de mastitis (Fernández *et al.* 2008, Mora

et al. 2015). Se ha estimado que el impacto económico de la mastitis puede rondar entre los \$60.000 y \$90.000 anuales en 100 vacas (Guimarães *et al.* 2017). Este impacto fue evaluado en granjas lecheras de Suecia por Nielsen *et al.* (2009), quienes mencionan que los casos de MC y MSC pueden alcanzar pérdidas económicas de €275 (USD\$308) y €60 (USD\$67), respectivamente. Por otro lado, Momena *et al.* (2017) han estimado que el costo de la mastitis varía entre USD\$64 y USD\$584 por vaca/año, con variaciones entre hatos.

Recientes estudios han demostrado que el principal daño de la mastitis se refleja en la producción de leche (Nielsen 2009, Guimarães *et al.* 2017), alcanzando hasta 75% de reducción de la producción, lo que representa un 54,9% de la pérdida económica total por mastitis (Guimarães *et al.* 2017). Por su parte, Huijps *et al.* (2008) anotan que la reducción de la producción representa el 100% de los costos en casos de MSC y de 36% para casos de MC. Aunado a esto, Nielsen (2009) reporta que la disminución diaria en la producción por vaca enferma (vacas con conteos mayor o igual a 500 000 CS/ml) se encuentra entre 0,7 Kg y 2,0 Kg para primíparas y 1,1 Kg y 3,7 Kg para multíparas, que en una lactancia de 305 días sumarían 150 Kg y 450 Kg para primíparas y multíparas, respectivamente.

Por su parte, un componente en el precio de la leche es el CCS, de manera que, al existir alta afectación por la mastitis, el precio de la leche se verá disminuido (Viguiet *et al.* 2009). En Costa Rica, la industria láctea maneja un sistema de bonificaciones por alta calidad de leche, brindando un aumento en el precio de la leche hasta de 4%, cuando los CCS son menores a 100 000 cs/ml. Se obtiene el nivel de precio base cuando el conteo alcanza las 500 000 cs/ml y se puede castigar el precio hasta un 25%, cuando la concentración de cs/ml es mayor a 1 000 000 (COOPROLE, 2016).

La necesidad de antibióticos para el control del agente causal es otro de los daños ocasionados por la mastitis; se utilizan unos 161 tratamientos al año para 100 vacas (Guimarães *et al.* 2017). Los costos de eliminación de la enfermedad pueden considerarse un rubro importante, representando hasta el 16% de las pérdidas económicas (Huijps *et al.* 2008). Sumado a eso, la leche que se obtiene de vacas enfermas con mastitis no puede ser utilizada para consumo humano, puesto que es

tratada con antibióticos y el bienestar animal puede verse comprometido si la enfermedad avanza (Gráf y Mikó 2015).

A pesar de que la mastitis es una enfermedad de la glándula mamaria, se han observado efectos en la reproducción de las vacas lecheras. Su impacto es similar al producido por el estrés calórico, afectando el eje hipotálamo-hipófisis-ovario, donde el principal efecto de la mastitis aguda y crónica se observa con un retraso en la producción preovulatoria de la hormona luteinizante (LH) y la ovulación retardada (Lavon *et al.* 2010, Roth y Wolfenson 2016). Por otra parte, Lavon *et al.* (2016) proponen que, una buena alternativa para aumentar la fertilidad en vacas lactantes con MSC es la utilización del protocolo Ovsynch seguido por la inseminación a tiempo fijo.

2.3 Métodos de detección de mastitis

Los métodos más tradicionales para la detección de mastitis subclínica han sido el conteo de células somáticas, la Prueba de California para Mastitis (CMT por sus siglas en inglés) y cultivo de bacterias (Viguiet *et al.* 2009, Digiovani *et al.* 2016). Adicionalmente, Viguiet *et al.* (2009) mencionan que la clave para desarrollar un buen método de detección de mastitis es entender bien su patogénesis.

De acuerdo con Pyörälä (2003), la leche debe ser controlada en tres niveles para la detección de mastitis y determinación de la calidad: i) evaluación individual de los animales, ii) prueba a gran escala para leche a granel y iii) comprobación en la planta de procesamiento. Existe una tendencia de transformación de los sistemas de ordeño para que sean más automatizados, por lo tanto, la detección utilizada en los sistemas tradicionales, como la Prueba de California y el conteo de células somáticas se hace poco eficiente y costoso. Según Momena *et al.* (2017) en los sistemas de ordeño automatizados, el contacto que tiene el personal con las vacas individuales es menor, esto implica que se deba de incurrir en métodos de detección en la línea de ordeño que sean de fácil medición, con niveles de sensibilidad, especificidad y precisión aceptables.

2.3.1 Conteo de células somáticas (CCS)

Este parámetro está relacionado con el estado sanitario de la ubre, conocida como una medida cuantitativa que muestra el total de células en leche, funcionando como un indicador de la calidad de ésta (Porporatto y Felipe, 2010; Guerrero-Quiceno *et al.* 2018).

De acuerdo con Porporatto y Felipe (2010), del total de células somáticas que se pueden encontrar en la leche de un cuarto enfermo, el 85% corresponden a glóbulos blancos, (60% son Macrófagos, 25% son Linfocitos y 15% son Neutrófilos); el restante 15% corresponde a células de descamación provenientes de los tejidos de la glándula mamaria. Estas células surgen como una respuesta humoral ante la infección por medio de agentes patógenos como lo son las bacterias de la mastitis.

Existen estándares con los cuales se clasifican los tipos de mastitis. Algunos autores reportan que conteos mayores a 200 000 cs/ml constituyen el umbral por encima del cual se presenta una mastitis subclínica (Kamphius *et al.* 2008, Sharma *et al.* 2011). Por otra parte, Bytyqi *et al.* (2010) sugiere que incluso conteos mayores a 100 000 cs/ml pueden ser considerados como indicadores de la presencia de mastitis subclínica. De manera similar, Guimarães *et al.* (2017) reportan 250 000 cs/ml como un umbral ideal para vacas de la raza Holstein.

Por sí mismo, el conteo de células somáticas ha sido utilizado para la determinación de cuadros de mastitis clínica y subclínica, y es considerada como una de las mediciones más aproximadas. No obstante, el valor predictivo de esta medición puede aumentar mediante la transformación logarítmica. En la literatura se reportan estudios con muchas variaciones para su aplicación por lo que se ha requerido una estandarizar mediante la definición de la variable score de células somáticas (SCCS por sus siglas en inglés) (Shook y Schutz 1994, Ankinakatte *et al.* 2013, Sharma *et al.* 2013). En 1994, Shook y Schutz demostraron una correlación positiva entre mastitis clínica y el SCCS, lo que hace a esta variable de gran importancia para la selección de los animales.

2.3.2 Conductividad eléctrica (CE)

El aumento en el conteo de células somáticas es proporcional a una mayor concentración de iones en disolución como: Na⁺, K⁺ y Cl⁻, de manera que está ligado a un incremento de la conductividad eléctrica (Ribeiro *et al.* 2016). Esta variable ha sido

utilizada recientemente como una forma rápida para la detección de mastitis con dispositivos portátiles (Ribeiro *et al.* 2016) y sistemas de ordeño automatizados (Momena *et al.* 2017). A pesar de utilizar la conductividad eléctrica como predictor de mastitis y es uno de los métodos más utilizados, es cuestionado por la obtención de resultados sumamente variables (Momena *et al.* 2017). Se ha encontrado evidencia de que los dispositivos que miden la conductividad eléctrica se aproximan a un 50% de detección de casos correctamente (Ribeiro *et al.* 2016). Mollenhorst *et al.* 2010 han reportado que estos resultados pueden mejorar cuando se analizan los cuartos de forma independiente y que la combinación de esta variable con la medición de CCS en línea puede mejorar aún más la tasa de éxito. Adicionalmente, Momena *et al.* (2017) mencionan que la incorporación de información extra como la producción de leche, el flujo de leche, el número de ordeños incompletos, pueden aumentar la precisión de la detección y la capacidad para determinar el inicio temprano de la mastitis.

Ribeiro *et al.* (2016) proponen la interpretación de los resultados de conductividad eléctrica absoluta (medida de cada cuarto) y conductividad eléctrica diferencial (entre los cuartos) para la detección de mastitis (Cuadro 1).

Cuadro 1. Interpretación de los valores de conductividad eléctrica absoluta (CEA) y la conductividad eléctrica diferencial (CED) en la leche.

CEA (mS/cm)	CED (mS/cm)	Interpretación
< 6,2	< 0,5	Leche normal
≥ 6,2	< 0,5	Leche anormal, presencia de calostros o final de lactación
≥ 6,2	≥ 0,5	Mastitis subclínica
< 6,2	≥ 0,5	Mastitis subclínica

Fuente: Ribeiro *et al.* (2016).

Kamphius *et al.* (2008) determinaron que la implementación de CCS en línea puede generar mejores resultados que la implementación de la CE, por lo que se obtienen tasas de éxitos mayores (de 12,7% a 15,6% y de 6,9% a 11%, respectivamente) y tasas de falsas alarmas menores (de 2,9 a 3,7 y de 4,7 a 7,8, respectivamente). Estos modelos fueron ajustados a una sensibilidad de 80%.

2.3.3 Sistemas en línea

El aumento del tamaño en los hatos lecheros y la necesidad de reducir los costos de operación han creado la necesidad de implementar los sistemas automatizados en línea, los cuales deben integrar herramientas de detección de MC y MSC. Los esfuerzos en este campo están enfocados en la búsqueda del mejor sistema de alarmas en presencia de episodios de mastitis. Algunos de los mecanismos que se integran a estos sistemas son la conductividad eléctrica (CE), los conteos de células somáticas en línea (BMSCC, por sus siglas en inglés), y los filtros integrados para la detección de cuadros clínicos; los últimos están limitados por la disposición de personal capacitado y el BMSCC puede ser un equipo muy costoso (Kamphius *et al.* 2016).

De acuerdo con Kamphius *et al.* 2016, el objetivo principal de los sistemas en línea se basa en la detección oportuna de los episodios de mastitis y a la vez obtener el mínimo posible de falsos positivos, de manera que se logre una reducción en costos de revisión de vacas con falsas alarmas.

2.3.4 Otros métodos

Considerando la sintomatología de la glándula mamaria, la inflamación de la ubre puede ser utilizada como un método de detección de MC. De este modo, se han desarrollado métodos como la medición de la firmeza de la ubre mediante un dinamómetro, esto solo se ha validado como una herramienta eficiente para la determinación de MC, resultando en una sensibilidad de 64,3% y especificidad: 89,7%, para vacas primíparas; y sensibilidad: 62,5% y especificidad de 96,7% para vacas multíparas (Rees *et al.* 2016).

2.4 Modelos alométricos para la detección de mastitis

La combinación de variables en los modelos para generar resultados de mayor confiabilidad en los sistemas de detección ha sido implementada por varios autores. Kamphuis *et al.* (2008) determinaron que la combinación de la conductividad eléctrica y la medición en línea del conteo de células somáticas pueden mejorar hasta dos o tres veces la tasa de éxito, conocido como valor predictivo positivo, y disminuir dos o tres veces la tasa de falsas alertas, si lo comparamos con el resultado obtenido cuando se utilizan estos parámetros como único parámetro de medición.

Estos enfoques implicaban el uso de los umbrales de la CE (rango de 7,5 mS/cm a 10 mS / cm), la prueba de más de 250 índices y un método de control estadístico de procesos. Los índices se desarrollaron mediante la combinación de factores (y niveles dentro de cada factor), tales como incremento medio de laminación condicional, porcentaje de variación, desviación media absoluta, la media de error, infectada con relación no infectado, todo en relación con la media móvil (3-9 puntos de datos), ya sea del cuarto afectado o la media de los cuatro trimestres (Momena *et al.* 2017).

2.5 Inteligencia artificial en la detección de mastitis

Las redes neuronales artificiales (RNA) se pueden considerar, junto con sus componentes, la unidad base de la inteligencia artificial, que a su vez son la base de la automatización de procedimientos que se asocian con el pensamiento humano, dentro de los cuales se pueden mencionar: el aprendizaje, la resolución de problemas y la toma de decisiones (Ferrero 2014).

Las redes neuronales artificiales intentan replicar el comportamiento del cerebro humano, y se caracteriza por el aprendizaje automático (*Machine learning*), utilizando la experiencia y el conocimiento que le proveen un conjunto de datos. Aunque esquemáticamente, las RNA y el cerebro, tienen una estructura neuronal similar, su aprendizaje proviene de fuentes diferentes. Las RNA pueden ser programadas y se alimentan de un conjunto de datos, mientras que las neuronas del cerebro captan información de su entorno que traducen en aprendizaje. En términos sencillos, una red neuronal se puede comprender como una forma flexible de extender la regresión lineal (López y Fernández, 2008). RNA se ha utilizado para analizar grandes y complejas bases de datos en explotaciones lecheras, de modo que se pueda hacer uso eficiente de los datos y elucidar información.

García (1999) señala que la información de una red neuronal es almacenada de manera asociativa en todos los nodos de la RNA, conocido también como memoria asociativa. La salida de la red es una respuesta de la asociación entre los nodos en relación con el peso brindado a cada uno. El autor explica que una memoria asociativa es una aplicación de dato a dato que permite recuperar la información, aún con elementos faltantes.

Por ejemplo, Ankinakatte *et al.* (2013) desarrollaron y compartieron dos métodos para la detección temprana de mastitis en ganado lechero en Tjele, Dinamarca. Estos autores utilizaron RNA y Modelos Generales Aditivos para predecir mastitis, basados en datos generados en línea. Ambos métodos probaron ser similares en cuanto a la habilidad de predecir mastitis en ganado lechero, además demostraron que incluir el SCCS en los modelos de predicción aumenta la habilidad de predicción en un 5%. Nielen *et al.* (1995), en uno de los primeros reportes sobre detección de mastitis a través del uso de redes neuronales artificiales sugirieron un método de predicción de mastitis temprana a través de la observación e interpretación de parámetros como la CE, temperatura de la leche, actividad de la vaca, entre otras. La investigación concluye que una herramienta de predicción con base en redes neuronales diferencia correctamente cuartos enfermos por mastitis de cuartos sanos de forma temprana. Estos autores usaron CE como variable predictiva en ganado bovino lechero de encaste, Holstein Friesian y Holandes Friesian en Ultrech, Holanda.

En el 2009, Hassan *et al.* usaron redes neuronales artificiales para detectar patógenos (agentes causales) de mastitis bovina. Estos autores obtuvieron porcentajes de especificidad que varían de 97%-80% en modelos de redes neuronales supervisados y de 99% en los modelos no supervisados. Estos autores, además, aseguran que el índice de resistencia eléctrica y el score de células somáticas son parámetros asociados con mastitis bovina, que se pueden utilizar para construir robustos modelos de detección temprana de mastitis en varios estados progresivos de la enfermedad.

Cavero *et al.* (2008) implementaron las redes neuronales para la detección temprana de mastitis en sistemas de ordeño automatizados, la presencia de mastitis fue definida por dos parámetros en dos tratamientos: i) presencia de tratamiento contra mastitis y ii) conteo de células somáticas arriba de 100 000 cs/ml o 400 000 cs/ml, según el tratamiento. Las variables utilizadas para el estudio como datos de entrada fueron la conductividad eléctrica, la tasa de producción de leche, la tasa de flujo de leche y los días en lactación. Establecieron la sensibilidad mínima en 80%, y obtuvieron resultados de especificidad entre 51,1% y 74,9% y una tasa de error entre 51,3% y 80,5%, respectivamente, para los dos tratamientos mencionados.

Los algoritmos de aprendizaje automatizado como el RNA han demostrado ser útiles en otras aplicaciones técnicas en el campo del ganado lechero como la fertilidad del hato. Shahinfar *et al.* en el 2014 en Waterton, Wisconsin demostraron que es posible predecir preñeces en vacas lecheras alimentando algoritmos de ML con datos de manejo e historial del animal, de modo que se reducen costos relacionados a reproducción.

El algoritmo Naive Bayes es un clasificador probabilístico simple que calcula un conjunto de probabilidades contando la frecuencia y las combinaciones de valores en un conjunto de datos determinado. El algoritmo usa el teorema de Bayes y asume que todos los atributos son independientes dado el valor de la variable clase. Esta suposición de independencia condicional, rara vez se cumple en situaciones del mundo real, por lo tanto, la categorización es un poco ingenua, no obstante, el algoritmo tiende a funcionar bien y aprende rápidamente en varios problemas de clasificación supervisada (Patil y Sherekar 2013)

2.6 Comparación de métodos de diagnóstico

Cuando se desarrollan métodos de detección o predicción de mastitis es necesario comparar estos métodos entre sí para entender de manera general su utilidad y eficacia. Los modelos de detección se pueden implementar como pruebas diagnósticas para clasificar vacas o cuartos lactantes como sanos o enfermos. En este estudio se incluyen dos términos epidemiológicos para comparar entre métodos de diagnóstico: sensibilidad y especificidad. El resultado del método de diagnóstico es una clasificación binaria, sano o enfermo si un cuarto sufre o no de mastitis (Kamphuis 2010).

2.6.1 Sensibilidad

Esta variable es la probabilidad de que un individuo sea detectado como enfermo, cuando realmente lo está. Para obtener el valor de la sensibilidad, se analizan los animales enfermos detectados correctamente o “verdaderos positivos” (VP) y los enfermos detectados como sanos o “falsos negativos” (FN), poniendo en términos de probabilidad cuál es el potencial del método de detección para diagnosticar una enfermedad correctamente.

2.6.2 Especificidad

Esta variable se conoce como la probabilidad de que un individuo pueda ser detectado como sano, cuando efectivamente lo está. Para obtener la especificidad, se analizan los animales sanos detectados correctamente o “verdaderos negativos” (VN) y los sanos detectados como enfermos o “falsos positivos” (FP), poniendo en términos de probabilidad cuál es el valor del método de detección para diagnosticar una enfermedad correctamente.

La sensibilidad y la especificidad son interdependientes y varían cambiando el umbral de diagnóstico de la enfermedad. Por ejemplo, si se incrementa el umbral de 200.000 CS/ml a 500.000 CS/ml, menos animales van a ser clasificados como enfermos. Del mismo modo, más animales van a ser diagnosticados como sanos, por ende, si la sensibilidad del método de diagnóstico disminuye, se puede esperar un aumento en la especificidad (Kamphuis 2010).

3 MATERIALES Y MÉTODOS

3.1 Descripción del sitio de estudio

El estudio se desarrolló en Finca Ganadera La Uno S.A, ubicada en La Marina de San Carlos, específicamente en las coordenadas CRTM05 458742.53 E y 1147645.93 N. Esta zona es caracterizada principalmente por la presencia de explotaciones ganaderas, mayoritariamente para la producción de leche.

El manejo del hato en producción se hace de forma intensiva, utilizando un estabulado de alrededor de 10 000 m², donde se ofrece una dieta balanceada que es revisada constantemente, de acuerdo con criterios productivos y reproductivos del hato con el propósito de optimizar la producción (Figura 1). El suministro de alimento obedece a un ciclo mecanizado desde el establecimiento, mantenimiento, cosecha, acarreo, mezcla y distribución de forraje, junto con alimentos balanceados para completar una ración total mezclada (RTM)



Figura 1. Infraestructura para la alimentación de las vacas en producción. Finca Ganadera La Uno S.A, San Carlos, 2017.

Además, la finca cuenta con 20 ha para el pastoreo de las vacas secas, zonas de producción de forrajes rastreros como estrella africana (*Cynodon nlenfluensis*) y erectos como sorgo (*Sorgum bicolor*) y maíz (*Zea mays*), zonas de sesteo (espacio con condiciones idóneas para el descanso de los animales) e infraestructura para la

alimentación diaria de los reemplazos y protección del hato ante las condiciones climáticas adversas. Para la limpieza y desinfección de las instalaciones, tanto galerones de de alimentación como sala de ordeño, se utiliza una solución de ácido peracético (LIPERAX) que se aplica hasta dos veces por semana.

En finca Ganadera La Uno se cuenta con un promedio anual de 111 vacas en ordeño y cerca de 70 reemplazos, que constituyen la base genética de las siguientes generaciones en producción. La finca promueve una buena distribución de partos en el año para mantener la producción constante, sin embargo, se observan fluctuaciones. Se cuenta principalmente con animales de la raza Holstein, Jersey, Pardo Suizo y sus cruces.

La sala de ordeño de la lechería es de tipo espina de pescado, con capacidad para 10 vacas por fila. Cada estación de ordeño cuenta con una consola individual que permite controlar el ordeño y monitorear algunos parametros relacionados con cada animal (Figura 2).



Figura 2. Vista general de la sala de ordeño en finca Ganadera La Uno S.A, San Carlos, 2017.

La finca es un modelo especializado de producción de leche donde se realizan dos ordeños diarios. El primero se inicia a las 3 am y el segundo a la 1 pm. La duración aproximada de cada ordeño es de una hora, sin embargo, esto depende de la cantidad

de vacas en el hato de ordeño. El tiempo de ordeño en cada vaca es de aproximadamente cuatro minutos.

El equipo de ordeño cuenta con un sistema automatizado (AMS, por sus siglas en inglés), que permite monitorear y controlar de manera eficiente el ordeño de cada vaca. El sistema monitorea la conductividad eléctrica (CE), la producción de leche (PL) y otros parámetros asociados a la reproducción. El sistema tiene una serie de alertas que indican si hay riesgo de que una vaca se encuentre enferma por las alteraciones en la CE y la producción láctea. Cuando estas variaciones son mayores a 12% (sube CE y baja PL) (Figura 3), el equipo indica que la vaca podría tener problemas de mastitis, de manera que debe ser apartada para evaluarse con más detalle, sea por CCS, CMT u otras pruebas más específicas, como el cultivo bacteriológico.



Figura 3. Pantalla de control de ordeño, producción de leche (Kg), conductividad eléctrica (mS/cm) y diferencia (%) respecto a ordeños anteriores. Finca Ganadera La Uno S.A, San Carlos, 2017.

3.1.1 Protocolo de ordeño

Durante el estudio, se siguió con el protocolo de ordeño de la finca, para provocar alteraciones mínimas en cuanto al estrés de las vacas por el muestreo. A continuación se detallan los pasos del protocolo de ordeño:

- Las vacas ingresan a la sala de ordeño en grupos de diez animales, por cada lado de la fosa, donde se ubican barras electromagnéticas que detectan el número del animal. Se verifica que el número de identificación coincida con la posición del animal, de no ser así, se corrige en el sistema.

- Los animales que muestren suciedad en la glándula mamaria son sometidos a un lavado previo con una solución de abundante agua con Gluconato de Clorhexidina (GCH). A la totalidad, se le aplica una solución desinfectante con base en Peróxido de hidrógeno al 0,5% (Peroxigard), seguido de una espera de 30 segundos para que surta efecto.
- Seguidamente, se realiza el “despunte” (Figura 4.a), que consiste en extraer los primeros tres o cuatro chorros de leche de cada cuarto productivo de la vaca, el cual se seca y se limpia con una toalla desechable.
- Se coloca la máquina de ordeño en forma manual y se procede al ordeño.
- Cuando el sistema detecta una disminución del flujo de la tubería, hace un corte de vacío para retirar la máquina de manera automática, con el fin de evitar el sobreordeño.
- Finalmente, se coloca un sellador de pezones con yodo al 0,25% para evitar el ingreso de patógenos en la glándula mamaria y prevenir cuadros de infección (Figura 4.b).

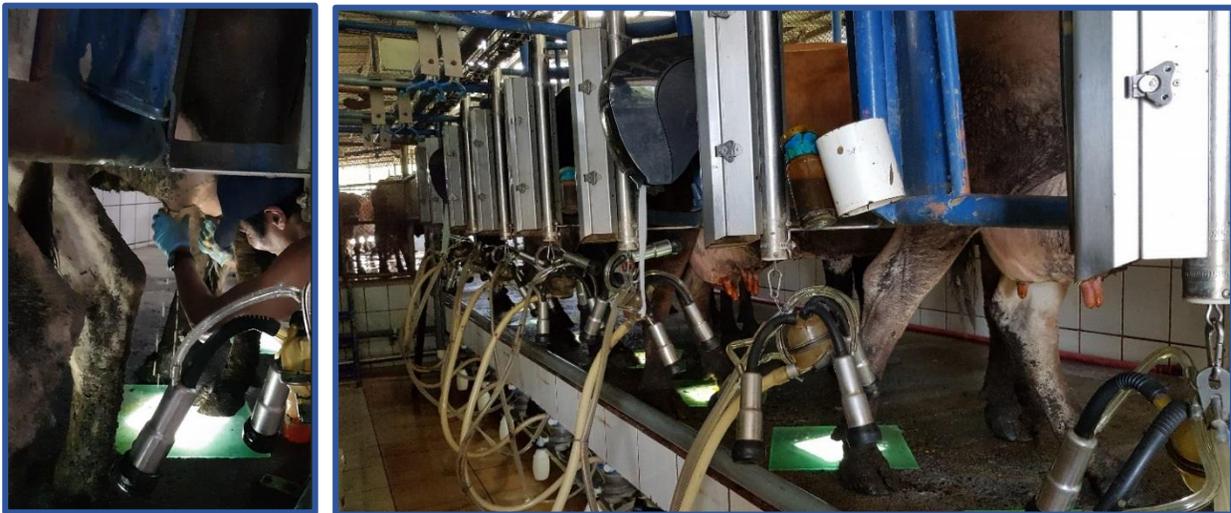


Figura 4. Protocolo de ordeño, a) despunte de la vaca: extraer primeros chorros, b) máquinas retiradas automáticamente y pezón recubierto con sellador. Finca Ganadera La Uno S.A, San Carlos, 2017.

3.1.2 Protocolo de lavado del equipo de ordeño

Al finalizar cada ordeño se sigue un protocolo de lavado de equipo, que se lleva a cabo con los siguientes pasos:

- Se conectan las pezoneras a la cañería de lavado para comenzar con el enjuague del equipo con agua recirculando a temperatura entre 40 °C y 50 °C.
- Se lava el equipo con un detergente alcalino (LACT), a una concentración de 2,64 ml por litro de agua y a una temperatura de 70 °C, durante 10 minutos.
- Se enjuaga el equipo con agua limpia a temperatura ambiente (25 °C), por cinco minutos.
- Seguidamente, con agua caliente se aplica el ácido removedor de piedra de leche (STONCIDE), a razón de 2,64 ml por cada litro de agua, durante cinco minutos.
- Antes del siguiente ordeño, se realiza un lavado con una solución de agua fría (25 °C) con hipoclorito de sodio (13%) a dosificación de 1,56 ml/l de agua durante cinco minutos.

3.2 Periodo de estudio

El estudio se desarrolló a partir del lunes 18 de setiembre del año 2017, con un muestreo semanal cada lunes durante once semanas, finalizando el lunes 28 de noviembre del 2017. Las muestras recogidas se enviaron al laboratorio de la Cooperativa de Productores de Leche Dos Pinos, y finalmente, se procedió con el análisis de éstos.

3.3 Material experimental

La población evaluada estaba compuesta por un hato de 111 vacas en producción, con edades que iban desde los 17 meses hasta los 154 meses, distribuidas en todas las etapas de lactación, desde el día 10 post parto hasta el final de la cosecha de leche. La cantidad de partos en el hato varió desde uno hasta ocho partos. Se obtuvieron muestras de vacas gestantes (n=403), abiertas (n=270), inseminadas (n=235), no gestantes (n=48) y estériles (n=6). El hato se compuso de animales de las razas Holstein, Jersey, Pardo Suizo y sus cruces. La producción promedio de estas vacas alcanzó los 14 kg por vaca por día.

3.4 Unidad muestral

Para este estudio se utilizó como unidad de medición toda vaca que ingresó a la sala de ordeño y que el equipo fue capaz de registrar los datos de producción de leche, conductividad eléctrica y, además se contó con el registro de la edad, días en lactación, número de partos y su estado reproductivo.

3.5 Variables regresoras y de respuesta

En el Cuadro 2 se presentan las variables regresoras y las variables de respuesta que fueron evaluadas en el estudio, con su respectiva abreviatura y equipo de medición. Para el caso del score de células somáticas (SCCS), se deriva de una transformación de CCS, con la ecuación utilizada por Sharma *et al.* (2013), de manera que se muestra la ecuación utilizada en el mismo cuadro.

Cuadro 2. Variables regresoras y de respuesta evaluados en el estudio de detección de mastitis subclínica, San Carlos, 2017.

Variable	Abreviatura	Unidad	Equipo
Regresora / Clasificación			
Conductividad eléctrica (máx.)	CE	mS/cm	SMARTFLO™
Producción de leche	PL	Kg/vaca/día	SMARTCONTROL™ METER – US
Edad	EDc	meses	Registros (SMARTDAIRY - US)
Días en lactación	DEL (DELC)*	días	Registros (SMARTDAIRY - US)
Número de partos	NP (NPc)*	n	Registros (SMARTDAIRY - US)
Estado reproductivo	ER	discreta	Registros (SMARTDAIRY - US)
Respuesta			
Conteo de células somáticas	CCS	cs/ml	Fossmatic Minor
Score de células somáticas	SCCS	--	$SCCS = \text{Log}_2(\text{CCS}/100\ 000) + 3^{***}$

*Nota: cuando una de las variables se categorizó para su análisis, se le agregó una "c" seguido de la abreviación

***CCS fue expresado en células somáticas por mililitro (cs/ml)

El estado reproductivo fue dividido en cinco categorías, que se definen a continuación, según el software utilizado por la finca y la numeración entre paréntesis representa el valor utilizado en el análisis estadístico. Las vacas abiertas (2) fueron aquellas que se encontraban desde el día uno después del parto hasta que se les realizó

la primera inseminación. A partir de este momento, se pasaron a llamar vacas inseminadas (3). Las vacas gestantes (1) fue el nombre asignado a la categoría de vacas a las que se les pudo comprobar preñez mediante palpación u otro método, las no gestantes (4) fueron aquellas que se comprobó, mediante el mismo método, la ausencia de preñez. Por último, la categoría de estéril (5) fue asignada a aquellas vacas que estaban destinadas a sacrificio, fuera por mastitis u otro motivo.

3.6 Categorías de estratificación del muestreo

Se hizo una categorización para el número de partos (NP), que se llamó número de partos categorizado (NPc), desglosados en: 1= un parto; 2= dos partos; 3= tres y cuatro partos; 4= cinco o más partos. Por otra parte, la edad categorizada (EDc) fue dividida en cinco categorías que se detallan de la siguiente forma: 1= menor o igual a 24 meses, 2= de 24,1 a 48 meses, 3= de 48,1 a 72 meses, 4= de 72,1 a 96 meses, y por último 5= 96,1 o más meses. Por último, para días en lactación (DEL), se crearon cinco niveles para la variable categorizada (DELc): 1= de 0 a 60 DEL, 2= de 61 a 120 DEL, 3= de 121 a 200 DEL, 4= de 201 a 300 DEL y 5= mayor a 300 DEL.

La variable de estado reproductivo (ER) de la vaca se utilizó como variable clasificatoria en el análisis de regresión lineal, para obtener modelos de acuerdo con la condición, sin embargo, estos modelos no mostraron resultados prometedores, de manera que se excluyeron de los resultados.

3.7 Metodología de toma de datos

Una vez ingresadas las vacas en producción a la sala de ordeño, cada vaca se posicionó en una de las 20 estaciones de ordeño (10 por cada lado). Se procedió con el muestreo de aquellas vacas que cumplían con las características antes mencionadas (acápite 3.4).

Las mediciones de conductividad eléctrica (CE) y producción de leche (PL) fueron tomadas automáticamente por el equipo. La CE fue monitoreada y registrada en el SMARTFLO™ (Figura 5.c.4) como CE máxima, es decir, la mayor CE de toda la leche que circuló por el sensor de CE. La producción de leche se obtuvo con el SMARTCONTROL™ METER – US (Figura 5.c.3). Estas mediciones fueron reflejadas

en la consola TOUCHPOINT™ (Figura 5.a), en tiempo real y se registran en la base de datos (BD).



Figura 5. a) Consola y retirado automático - TOUCHPOINT™ para control de ordeño, visualización e ingreso de información individual; b) vista general del equipo automático de ordeño; c) máquina de ordeño (1), frasco recolector (2), flujómetro – SMARTCONTROL™ METER - US (3), medidor de CE - SMARTFLO™ (4) y cable de retirado automático (5). Finca Ganadera La Uno S.A, San Carlos, 2017.

Utilizando la identificación de las vacas, se accedió al registro histórico, de manera que se extrajo las variables técnicas necesarias para el estudio (Cuadro 2) al término de cada ordeño en formato .xlsx, a excepción del CCS, que fue brindado por el laboratorio en formato .pdf, y que luego incorporado a la base de datos general.

Sin interrumpir el proceso normal de ordeño (acápite 3.1.1), una vez que las vacas estuvieron bien ubicadas en cada estación y debidamente identificadas, se le colocó una bolsa para muestras de leche, también identificada (ver Figura 6.a) junto al recolector (Figura 5.c.2). La muestra que se obtuvo fue representativa del ordeño de cada vaca, ya que es el resultado de una fracción separada en la cañería de leche que viene de la máquina de ordeño. El volumen aproximado que se requirió para enviar las muestras al laboratorio fue de 70 ml de leche por vaca.

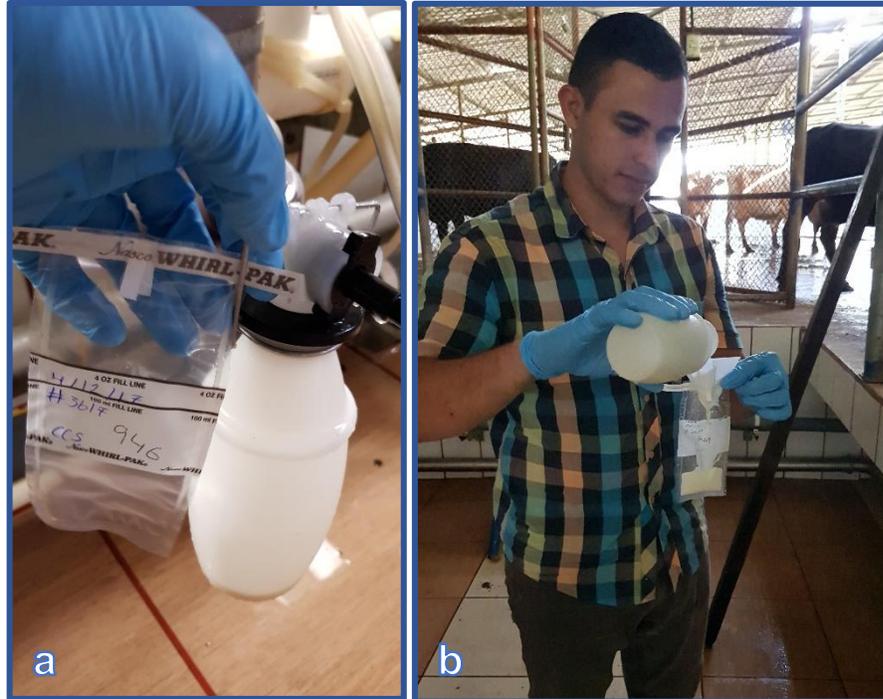


Figura 6. a) Identificación y colocación de la bolsa para muestreos de leche en el frasco recolector de la muestra, b) filtrado y vaciado de la leche en la bolsa de muestreo. Finca Ganadera La Uno S.A, San Carlos, 2017.

Para la identificación de estas muestras individuales se utilizó la fecha de muestreo, el número de vaca, el número de finca y el tipo de análisis (CCS) (Figura 6.a). Una vez que las bolsas quedaron debidamente selladas, se introdujeron en una hielera para bajar la temperatura y evitar la proliferación de microorganismo, por ende, la descomposición de la leche. Inmediatamente después de completar el ordeño, las muestras fueron llevadas al Laboratorio de Calidad de Leche Dos Pinos, ubicado en Ciudad Quesada, donde se le realizó el conteo celular somático. Los resultados fueron brindados transcurridos tres días después de cada muestreo. Con la base de datos completa, se calculó el SCCS utilizando la siguiente ecuación que se muestra en el Cuadro 2, donde CCS representa el conteo celular somático y mientras SCCS es el score de células somáticas, mediante la transformación con la fórmula indicada.

3.8 Análisis de datos

Una vez recopilada y tabulada la base de datos (n=966), se sometió a un proceso de depuración en el cual se eliminaron los registros incompletos y aquellos que mostraron extremos o valores atípicos (“outliers”). De la base de datos depurada (n=942), se extrajo al azar un 16,6% de los registros, los cuales se utilizaron para ejecutar la validación de los modelos elaborados. Se utilizó el programa estadístico InfoStat/P para la elaboración de los modelos alométricos o de regresión (Di Rienzo *et al.* 2017). Además, se utilizaron librerías de redes neuronales artificiales que se describen en el acápite 3.8.2 para hacer las predicciones con aprendizaje automático (*Machine Learning*). Por último, se implementó el programa estadístico R (R Core Team, 2018) para la elaboración de los gráficos con el paquete ggplot2 (Wickham, 2016).

3.8.1 Análisis exploratorio

En primera instancia se hizo una descripción de la población a nivel productivo, sanitario y económico. Se hicieron relaciones entre la PL y el CCS en función de los muestreos, también se comparó el ingreso económico bruto en función del SCCS. En este ingreso se contempló tanto el efecto del rendimiento de leche, como el precio de la leche afectado por los conteos de células somáticas (Anexo 4). Se definió el precio base de la leche en 320 colones por kilogramo.

Se hizo un análisis de componentes principales para determinar las correlaciones entre SCCS y las variables regresoras. Seguidamente, se sometieron a comparación las vacas sanas (0) y enfermas (1), según la PL, CE y CCS. También se comparó el SCCS, la PL y la CE en función de tres variables categorizadas (NPc, DELc y EDc). La prueba de comparación utilizada fue la de Tukey.

3.8.2 Modelos alométricos (MR)

Con los registros restantes (83,4% de la base de datos original), se establecieron asociaciones entre las variables independientes y la variable dependiente (SCCS), con el propósito de seleccionar las variables con mayor asociación. Con este set de variables se procedió a elaborar modelos de predicción para SCCS, en función de las variables predictivas. Los modelos obtenidos fueron lineales y linealizados, múltiples, con variables originales y transformadas. Posteriormente, se seleccionaron los modelos que

cumplieran con los mejores criterios de bondad de ajuste, (menor R^2 , más adecuada distribución de los residuos) criterios de información de Akaike (AIC) y bayesiano (BIC). Tanto los criterios de información (AIC y BIC) como el R^2 se tomaron del cálculo del programa estadístico. Finalmente, se validaron los modelos con el 16,6% de los datos seleccionado para este fin, se les aplicó la prueba ajuste Wilcoxon en el mismo programa y el sesgo se calculó con la siguiente ecuación.

$$S = \frac{\sum(\text{observados} - \text{estimados})}{\text{observados}} \times 100$$

3.8.3 Algoritmos de aprendizaje automático (*Machine learning*)

Los algoritmos de aprendizaje automático (ML por sus siglas en inglés) se entrenaron con el mismo set de datos que los modelos de regresión. Estos utilizan redes neuronales artificiales (RNA), aplicando el método de aprendizaje automático supervisado. Para correr el algoritmo, fue necesario añadir preguntas denominadas “características” y respuestas denominadas “etiquetas”, de manera que una vez que el algoritmo las combina, puede hacer la predicción.

En este caso, las “características” fueron las variables regresoras (independientes), que se utilizaron de forma simple o categorizada. Como “etiqueta” se utilizó la variable SCCS, ya que fue la variable de respuesta (dependiente), con mejores resultados en todos los modelos debido a su normalidad (Anexo 1 y Anexo 2).

El programa utilizado para entrenar y probar los algoritmos de *machine learning* fue Python mediante la librería Scikit-learn, de donde se obtuvieron también los algoritmos de aprendizaje supervisado de tipo clasificación (MLC) propuesto por Naive Bayes y la función “*LinearRegresion*” para el modelo de regresión que implementó *machine learning* (Pedregosa *et al.* 2011). Los archivos de Excel fueron convertidos a formato .csv y manipulados con el programa Pandas (Augspurger *et al.* 2019). En el Anexo 7 y el Anexo 8 se puede observar la programación realizada para entrenar al modelo de clasificación y regresión lineal, respectivamente.

Se probaron dos tipos de algoritmos de ML: i) clasificación (MLC) y ii) regresión (MLRG). En el caso de MLC, el algoritmo detectó patrones que intentaron clasificar los animales en dos grupos: 0) sanas y 1) enfermas (se definieron en el acápite 3.8.4).

Dentro de la categoría de MLC se entrenaron dos modelos, el primero utiliza como características las cuatro variables discretas o discretizadas (EDc, DELc, NPc y ER) y dos variables continuas (PL y CE) para predecir el estado de salud de la vaca, a este se le llamó MLCD. El otro modelo que se entrenó se le llamó MLCC, y contemplaba todas las variables continuas (ED, DEL, NP, PL y CE), ER se excluyó por ser una variable discreta. Al igual que MLCD, MLCC ubica el resultado en forma binaria (0,1).

Por otro lado, el de tipo regresión (RG) generó una salida de datos continuos de SCCS, que después fueron clasificados sanas y enfermas como se describe en el acápite 3.8.4. A este tipo de modelo se le llamó MLRG y se definió con la siguiente ecuación.

$$SCCS_{MLRG} = b_0 + m_0 * ED + m_1 * DEL + m_2 * NP + m_3 * PL + m_4 * CE$$

Por medio del entrenamiento del algoritmo con las reiteradas corridas de datos, se obtuvieron los valores de b_0 , m_0 , m_1 ,... m_4 . Posteriormente, estos se sustituyeron en la ecuación para calcular la predicción.

Todos los modelos de ML se pusieron a prueba con la base de datos de validación (n=157). De estos, solo para el MLRG se hizo la prueba de ajuste y sesgo anteriormente mencionados.

3.8.4 Sensibilidad y especificidad

La sensibilidad y la especificidad son datos generados para representar en porcentaje la capacidad de un método de predecir una condición determinada. En este caso, diferenciar entre vacas sanas y enfermas. Después de la validación, se hizo una evaluación del resultado obtenido, considerando criterios de sensibilidad (SEN), especificidad (ESP), tasa de error y precisión. Utilizando estos parámetros, fue posible comparar entre los modelos obtenidos y con otros modelos y métodos de detección de mastitis reportados en la literatura.

En la Figura 7 se representa la forma en que los datos fueron categorizados, de acuerdo con el resultado obtenido en el análisis de laboratorio (medición real) y la predicción del modelo alométrico y/o el algoritmo de aprendizaje a prueba.

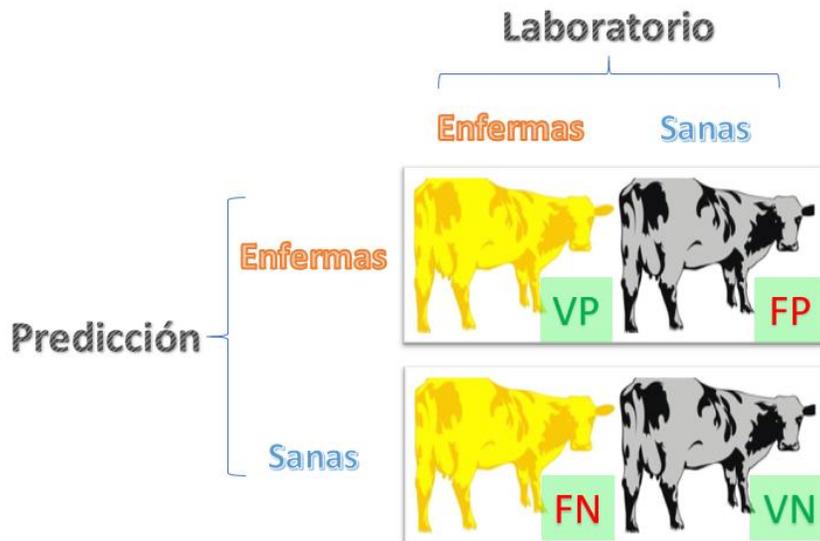


Figura 7. Relación entre el resultado del modelo y análisis de laboratorio para la obtención de parámetros de comparación (SEN, ESP, precisión y tasa de error).

Para utilizar estos parámetros de comparación, se determinó que el punto crítico (P_c) entre una vaca sana y una enferma fue de 200 000 cs/ml o un SCCS de 4. De esta manera, una vaca se consideró enferma cuando el conteo celular somático superó las 200 000 cs/ml ($SCCS > 4$). Cuando el resultado fue menor o igual a este P_c se consideró sana. Por tanto, como se observa en la Figura 7, cuando el modelo indicó que una vaca es sana y el resultado de análisis en laboratorio también es sano, se consideró que fue un “verdadero negativo” (VN); si el modelo indicó que la vaca es sana y el resultado del laboratorio es positivo (> 200000 cs/ml), se le llamó “falso negativo” (FN). Por el contrario, cuando el modelo predijo que las vacas estaban enfermas (positivas a mastitis), se consideró “verdadero positivo” (VP) cuando el resultado de laboratorio fue positivo, se determinó como “falso positivo” (FP), cuando el resultado de laboratorio fue negativo. El cálculo de SEN, ESP, precisión y tasa de error se calcularon de acuerdo con las siguientes ecuaciones.

$$SEN (\%) = \frac{VP}{(VP + FN)} \times 100$$

$$ESP (\%) = \frac{VN}{(VN + FP)} \times 100$$

$$Precisión (\%) = \frac{(VN + VP)}{(VN + FP + FN + VP)} \times 100$$

$$Tasa de error (\%) = \frac{FP}{(FP + VP)} \times 100$$

En el caso de los dos algoritmos de *machine learning* de tipo MLC (MLCC y MLCD), se hizo una preclasificación del resultado en sana (0) y enferma (1) para entrenar y obtener el resultado del modelo en esta forma. Las curvas de SEN y ESP fueron utilizadas para determinar gráficamente el punto de corte entre ambas curvas, conocido como *threshold*. También se calculó, a partir de la gráfica, el valor de ESP a una SEN fijada en 80%, como parámetro de comparación entre modelos. Asimismo, se crearon curvas ROC con el fin de determinar el potencial del modelo para el diagnóstico de la enfermedad, medido como el área bajo la curva ROC.

4 RESULTADOS Y DISCUSIÓN

4.1 Descripción de la población

En el Cuadro 3 se muestra el comportamiento general de la población evaluada respecto al conteo de células somáticas (CCS), además evidencia el alto grado de variabilidad que existe en los datos observados dada su desviación estándar. En términos generales, se observó que la población de vacas en producción de Finca Ganadera La Uno S.A mantuvo niveles promedio entre 391 000 cs/ml y 644 000 cs/ml durante el periodo en estudio, mostrando los conteos más altos en los muestreos uno y dos (511 000 y 644 000, respectivamente); posteriormente se mantuvieron entre 391 000 y 488 000 cs/ml. También se muestra que aproximadamente el 50% de las vacas se encontraban sanas (<200 000 cs/ml) y el otro 50% estaban enfermas (>200 000 cs/ml), con variaciones entre los muestreos. Esta característica del conjunto de datos permitió tener una cantidad de datos similar, por encima y por debajo del punto crítico de diagnóstico de mastitis (200 000 cs/ml), lo que favoreció el posterior análisis de sensibilidad y especificidad.

Cuadro 3. Comportamiento general del conteo celular somáticas (CCS) según muestreos semanales, durante el estudio. Finca Ganadera La Uno S.A, San Carlos, 2017.

Semana de muestreo	Promedio	Desviación estándar	Mediana	Mínimo	Máximo	Sanas	Enfermas
						CCS (cs/ml)	
1	511 351	840 583	176 500	28 000	5 061 000	54,1	45,9
2	644 567	906 498	284 000	25 000	4 719 000	41,1	58,9
3	488 225	642 602	198 000	10 000	3 862 000	50,6	49,4
4	435 461	556 423	233 000	20 000	2 483 000	44,9	55,1
5	420 632	533 832	190 000	30 000	2 881 000	51,7	48,3
6	393 933	519 163	166 000	21 000	2 843 000	56,2	43,8
7	412 022	508 179	191 000	10 000	2 410 000	52,7	47,3
8	476 604	748 193	225 000	22 000	5 477 000	47,3	52,7
9	386 148	538 917	175 000	9 000	2 817 000	53,4	46,6
10	473 307	801 897	195 000	25 000	5 878 000	52,3	47,7
11	391 310	461 544	210 500	24 000	2 291 000	47,6	52,4

Es importante destacar que la finca maneja una carga de antibióticos para el tratamiento de mastitis relativamente baja (se utiliza solamente en casos muy severos). De acuerdo con el productor, es más favorable, optimizar el sistema inmunológico mediante la aplicación de productos que estimulen el sistema linfático para la producción de células de defensa.

No se encontraron diferencias significativas entre las semanas de muestreo ($p > 0,05$), sin embargo, se observó una tendencia en el aumento de la producción conforme descendió el conteo celular somático del hato, tendencia que coincide con otros autores (Nielsen 2009, Guimarães *et al.* 2017, Gonçalves *et al.* 2018b). Adriaens *et al.* (2018) respalda esta información y añade que la disminución de producción se da tanto en el(los) cuarto (s) afectados de manera directa, como en los cuartos adyacentes de forma indirecta. Estos autores indican que esto se puede atribuir a la enfermedad sistémica de la vaca reflejada en la pérdida de apetito y necesidad de energía para la reacción inflamatoria. Por su parte, Green *et al.* (2006) mencionan que esta relación entre PL y CCS se da en gran medida, por un efecto de dilución y que debe ser estudiado con mayor profundidad, ya que la merma de la producción láctea puede estar sobreestimada.

4.1.1 Impactos económicos sobre la explotación

De acuerdo con varios autores los conteos altos de células somáticas por vaca afectan tanto la producción como la calidad de la leche (Nielsen 2009, Vásquez *et al.* 2014, Guimarães *et al.* 2017). En la Figura 8 se muestra la tendencia del ingreso económico bruto en función del score de células somáticas. En consecuencia, a lo que mencionan estos autores, se observó un doble efecto negativo al contar con niveles elevados de CCS en la leche, repercutiendo en el precio y en la rentabilidad de la actividad. Se utilizó SCCS ya que se normaliza distribución de las observaciones en comparación al CCS (Anexo 1).

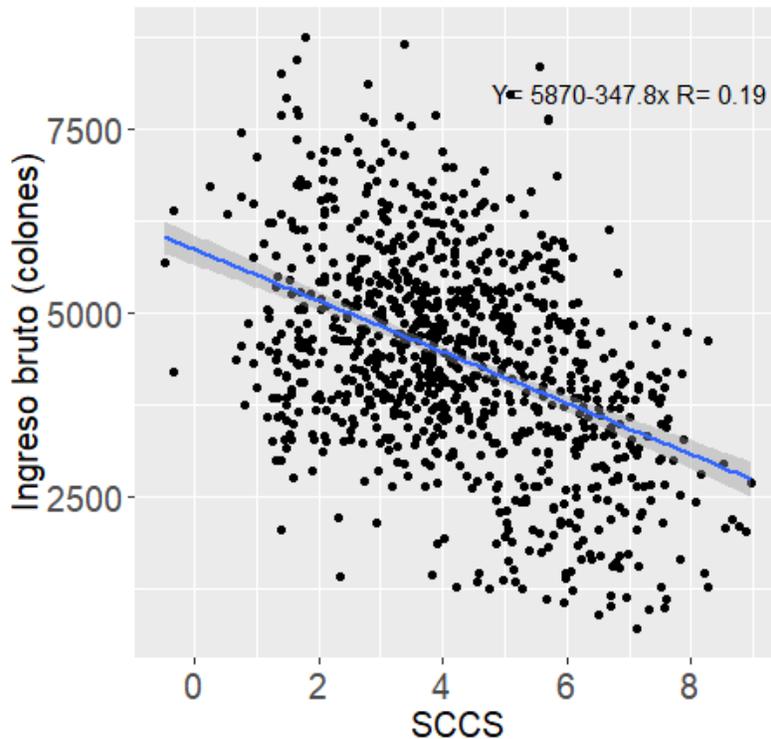


Figura 8. Comportamiento del ingreso bruto (colones) por vaca por día según score de células somáticas (SCCS) del hato en estudio. Finca Ganadera La Uno S.A, San Carlos, 2017.

Como se observa en la Figura 8, la tendencia del ingreso por vaca es menor conforme aumenta el SCCS. Poniendo en contraste dos vacas, una en el límite de las vacas sanas (SCCS= 4; CCS= 200 000 cs/ml) y otra vaca enferma (SCCS= 6; CCS= 800 000 cs/ml); se sustituyen los valores de SCCS en la ecuación de regresión obtenida, de manera que se obtiene un ingreso de 4478,8 colones ($x=4$) y 3783,2 colones ($x=6$), respectivamente. La diferencia de ingreso obtenida es de 695,6 colones por vaca por día, esto en una lactancia normal de 305 días puede acumular hasta 212 158 colones por vaca por lactancia. Visto de otra manera, si se considera esta cantidad en un hato de 111 vacas (caso de la finca en estudio), semanalmente representa un ingreso bruto extra de 486 920 colones. Al comparar las vacas sanas y las enfermas, se obtuvieron pérdidas en el ingreso promedio de un 22,4%. Resultados similares se ha reportado en México para vacas de la raza Holstein; estos autores reportan una reducción de ingresos de 16 y 24,5% con la presencia de mastitis subclínica (Arturo *et al.* 2014).

Partiendo de que SCCS es una transformación de CCS, lo obtenido refleja el efecto negativo de la mastitis sobre el ingreso económico de la actividad. Además del efecto, sobre la disminución del ingreso, por la caída en la producción y la disminución de la calidad de leche, otros autores mencionan pérdidas adicionales por el uso de antibióticos, descarte de leche e incluso se mencionan efectos negativos sobre la reproducción de los animales (Lavon *et al.* 2010, Gráf y Mikó 2015, Guimarães *et al.* 2017).

4.2 Relación entre las variables evaluadas

De acuerdo con el análisis de componentes principales (ACP) realizado con la totalidad de las variables evaluadas (biplot en la Figura 9a) fue posible explicar el 67,7% de la variación total. En este biplot se observa que la conductividad eléctrica (CE) resultó ser de menor importancia, además de estar positivamente correlacionada con el número de partos (NP). Lo anterior se corrobora con el biplot de la Figura 9b, el cual demuestra que, al excluir dicha variable, se obtuvo una ganancia en la explicación de la variación total (76,7%). Sin embargo, en el presente estudio no se consideró pertinente eliminar la conductividad eléctrica, puesto que la misma se ha considerado importante en otros estudios (Kaşıkçı *et al.* 2012, Ankinakatte *et al.* 2013) para predecir el score de células somáticas o en conjunto con este para diagnosticar la mastitis.

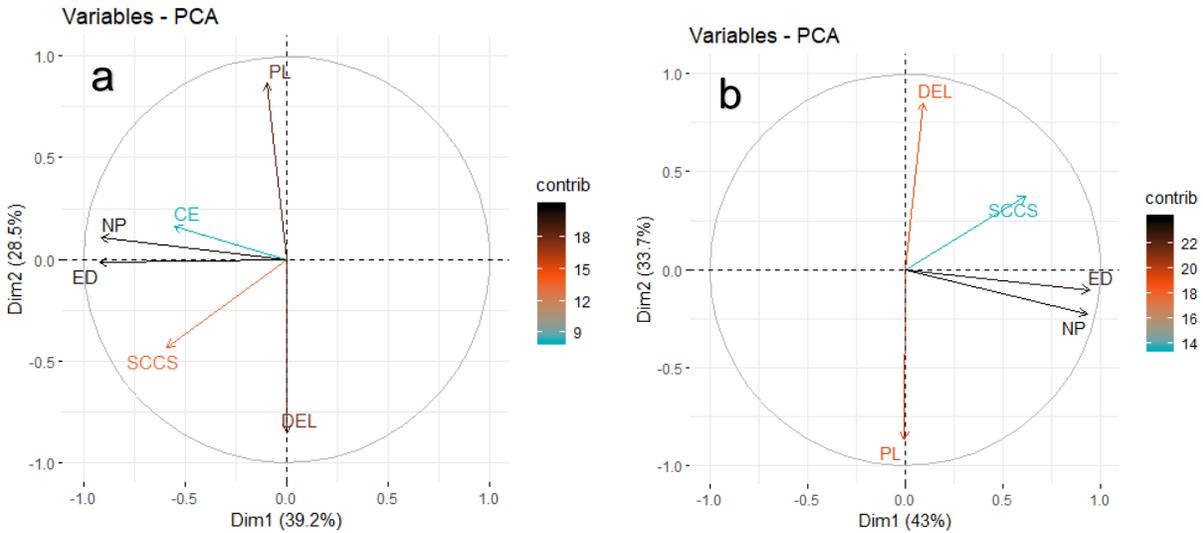


Figura 9. Análisis de componentes principales (ACP) para las variables regresoras y dependientes. a) considera la variable CE, b) excluye la variable CE. Finca Ganadera la Uno S.A, San Carlos, 2017.

4.3 Comparación entre las variables de estudio

En el estudio se encontró diferencias significativas entre vacas sanas y enfermas para CE, PL y CCS. Estas diferencias se aprecian en el Cuadro 4. En las vacas sanas se obtuvo una CE de 6,41 mS/cm, esto difiere de los obtenido por Norberg *et al.* (2004), quienes reportan una media de 4,87 mS/cm para vacas sanas, valor similar al presentado por Kaşikçi *et al.* (2012) quienes mencionan un rango de 4,0 mS/cm a 5,5 mS/cm para una leche normal (sana) a 25 °C. También se reportan niveles de CE para vacas con mastitis subclínica y clínica que corresponden a 5,37 mS/cm y 6,44 mS/cm, respectivamente (Norberg *et al.* 2004). Los resultados obtenidos sugieren que pueden existir diferencias entre hatos e incluso entre animales del mismo hato para niveles de CE, lo que hace difícil la estandarización de un método de detección de mastitis basado en esta variable. Esto hace necesaria la generación, estudio y validación de los modelos de predicción de mastitis basados en CE o que la incluyen.

Cuadro 4. Comparación de la media y desviación estándar de la CE, PL y CCS entre vacas sanas (0) y enfermas (1) ($p < 0.05$). Finca La Uno S.A, San Carlos, 2017.

Variable	Condición	Promedio	D.E.	Tukey
CE (mS/cm)	0	6,41	0,76	a
	1	6,79	1,00	b
PL (kg/día)	0	14,74	3,69	b
	1	13,21	4,33	a
CCS (cs/ml)	0	99 975	51 896	a
	1	816 077	778 048	b

***Letras diferentes indican diferencias significativas ($\alpha=0,05$), de acuerdo con la prueba de comparación Tukey.

En otro orden la Figura 10, muestra el análisis de comparación entre las variables categorizadas y las continuas. Las variables continuas utilizadas fueron el SCCS, CE y PL y las variables categorizadas fueron número de partos, días en lactación y edad. En respuesta se obtuvo una tendencia de SCCS creciente en todas las variables de clasificación. En primer lugar, se observa que el SCCS aumentó según el número de partos (Figura 10a), esto sugiere que una vaca tendió a aumentar los problemas de mastitis conforme tiene más partos. Esta información coincide con lo mencionado por Mora *et al.* (2015), quienes indican que el número de partos tiene un efecto significativo en la incidencia de mastitis. Solamente las vacas primíparas mostraron un promedio de SCCS bajo el límite de las vacas sanas (<4), el equivalente en términos de CCS para el promedio de la categoría 1 de partos es de 101 000 cs/ml, conteo similar al obtenido por Zambrano y de Pinho (2008) para vacas Girolando de primer parto hasta 153 días de lactación. Adicionalmente, Niozas *et al.* (2019) respaldan el resultado obtenido, e indican que las vacas primíparas tienen menores conteos de células somáticas en comparación a las vacas que tienen 2 o más partos.

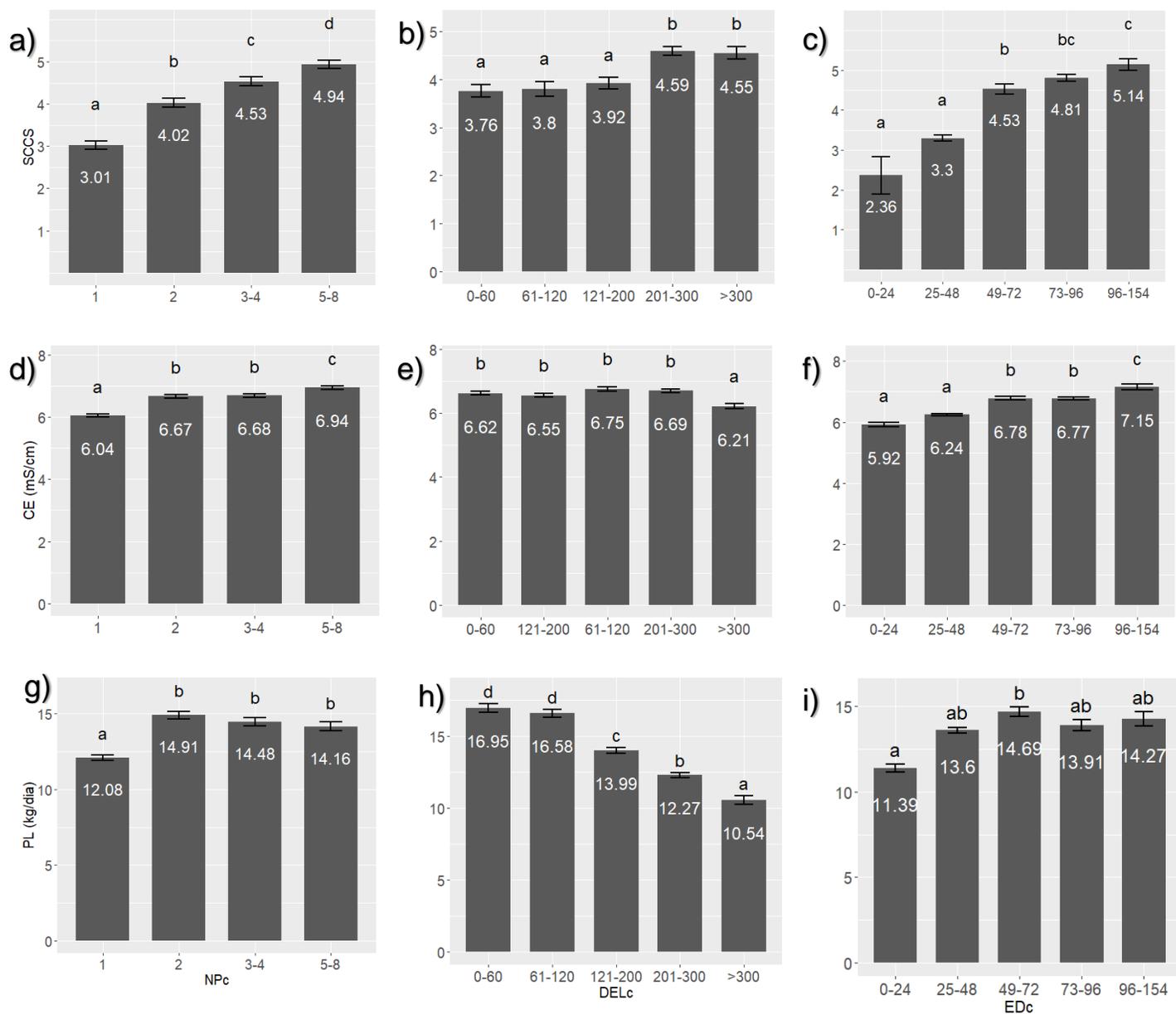


Figura 10. Comportamiento de tres variables continuas (score de células somáticas: SCCS, conductividad eléctrica: CE, y producción de leche: PL) según las variables categóricas (número de partos: NPc, días en lactación: DELc y Edad: EDc). a) SCCS vsr NPc, b) SCCS vsr DELc, c) SCCS vsr EDc, d) CE vsr NPc, e) CE vsr DELc, f) CE vsr EDc, g) PL vsr NPc, h) PL vsr DELc, i) PL vsr EDc. Finca Ganadera La Uno S.A, San Carlos, 2017. ***

***Letras diferentes indican diferencias significativas ($\alpha=0,05$), de acuerdo con la prueba de comparación Tukey.

En cuanto a los DELc, SCCS se comportó similar en las primeras tres categorías, es decir, hasta los 200 días de lactancia; a partir de los 201 días subió el score de células somáticas significativamente ($p < 0.05$) hasta el final de la lactación. Niozas *et al.* (2019) reportan un efecto significativo ($p < 0.0001$) de los días en lactación con el aumento del conteo celular somático, lo cual coincide con la tendencia de la Figura 10b y es respaldado por Sharma *et al.* (2013), quienes indican que las vacas con lactancias extendidas tienen más riesgo de incidencia de mastitis. Hertl *et al.* (2018), revela que el aumento en los casos de mastitis clínica durante los primeros 100 días de la primera lactancia aumenta la tasa de aparición de nuevos casos de MC en lactaciones posteriores y, por ende, la tasa de sacrificio de estos animales aumenta un 34% en cada caso de MC durante este periodo.

Se observó un comportamiento similar entre NPc y EDc cuando se compararon con el SCCS, esto se debe a que estas dos variables están altamente correlacionadas ($R^2 = 0,90$). Otros autores han determinado que la edad (EDc) tiene un efecto significativo en el SCCS, encontraron diferencias en un intervalo de confianza del 95%. Esto puede atribuirse, según Sharma *et al.* (2013), a la fisiología de la ubre que hace que la glándula mamaria sea más propensa a la mastitis. El aumento de SCCS podría atribuirse a un aumento de la incidencia de mastitis, ya que el número de partos y la etapa en lactancia no tiene efecto significativo en el SCCS cuando los animales son bacteriológicamente negativos (Niozas *et al.* 2019).

La CE de la leche está determinada por la concentración de aniones y cationes en ella, los iones Na^+ , K^+ y Cl^- aumentan la CE de la leche (Kaşikçi *et al.* 2012). Recientemente se han reportado estudios que utilizan la CE como un predictor de mastitis (Elizalde *et al.* 2009, Ribeiro *et al.* 2016, Momena *et al.* 2017). Norberg *et al.* (2004) reportan que la diferencia de CE inter-cuartos aporta más para la predicción de mastitis que la medición general, sin embargo, esta medición puede resultar más costosa si no se cuenta con el equipo adecuado. Además, sugiere el estudio de parámetros genéticos para mejorar este estimador.

De los estudios mencionados, ninguno ha reportado diferencias de CE para número de partos, días en lactación o edad. En la Figura 10 (d,e y f) se muestra la relación NPc,

DELc y EDc con CE, respectivamente. Las diferencias más evidentes se obtuvieron para NPc y EDc, en la variable DELc, sólo se encontraron diferencias ($p < 0,05$) en las vacas con más de 300 días en lactación, con menor presencia de iones en la leche. En cuanto a EDc, las vacas que presentaron mayor CE fueron las que tenían más de 96 meses de edad, al contrario de esto, las vacas más jóvenes (17-24 meses) fueron las que presentaron menor CE. La tendencia en NPc es similar a EDc, debido a que estas dos variables están altamente correlacionadas ($R^2=0,90$). Las vacas con un parto obtuvieron la CE más baja; dos, tres y cuatro partos, se comportaron similares, y por último, las vacas con 5 o más partos fueron los animales con más presencia de iones en la leche.

La producción de leche (PL) para cada categoría de parto se puede ver en la Figura 10g. Se encontraron diferencias significativas entre las vacas con un parto (12,08 kg/día) y las vacas de dos o más partos (>14,16 kg/día) ($p < 0,05$), no hubo diferencias entre las vacas de dos partos, tres y cuatro partos y cinco o más partos ($p > 0,05$). Varios autores muestran resultados similares, indicando que la primera lactancia tiene significativamente menor producción en comparación a dos o más lactancias. De acuerdo con la literatura los mayores rendimientos se encuentran entre tercera y quinta lactancia (Carvajal-Hernández *et al.* 2002, Niozas *et al.* 2019). Los resultados obtenidos sugieren un aumento significativo desde la segunda lactancia (Figura 10g).

Los DEL representan los días que han transcurrido después del parto de una vaca. De este modo, al categorizarse (DELc) estos grupos se pueden asociar a ciertos eventos fisiológicos, que afectan de alguna manera la producción y la reproducción. Los resultados muestran diferencias significativas ($p < 0,05$) para cada etapa de lactación (DELc), con excepción de las primeras dos etapas, donde se obtuvieron los rendimientos más elevados antes de los 121 días post-parto (Figura 10h), sin observarse diferencias entre la primera categoría (0-60 días) y la segunda (61-12 días) ($p > 0,05$), además, en las categorías siguientes sucesivamente fueron decreciendo en producción con diferencias significativas ($p < 0,05$). De acuerdo con Gonçalves *et al.* (2018a) las menores pérdidas de producción de leche se obtienen en el pico de lactación, además mencionan un aumento de las pérdidas en rendimiento a mayor avance en los DEL.

4.4 Selección del modelo

4.4.1 Modelos alométricos

En el análisis de regresión se obtuvieron cuatro modelos lineales o linealizados que se observan en el Cuadro 5. Todos fueron significativos, tanto para los coeficientes como para las variables utilizadas en cada modelo. De estos cuatro, los que presentaron mejores resultados como un alto coeficiente de determinación ajustado (R^2), menor AIC y menor BIC fueron MR3 y MR4. No obstante, se encontró que el modelo MR3 fue el que presentó mayor sesgo (7,61%). Aun así, se determinó que el sesgo de todos los modelos fue alto (>2%) y los R^2 bajos (<0.40). Todos los modelos se caracterizaron por utilizar variables continuas y discretas (o discretizadas), además se observó un efecto positivo sobre los criterios de selección en los modelos MR3 y MR4, aunque con un aumento en la complejidad del modelo, debido a la utilización de un mayor número de variables y a la combinación entre ellas, lo que pudo propiciar mayor absorción de variabilidad cuando los individuos se agruparon en conjuntos. Este efecto positivo ocurrió cuando se trató de parámetros de ajuste lineal, no así, para análisis de SEN y ESP.

Cuadro 5. Comparación de criterios de selección de modelos de regresión (MR) y sus respectivas ecuaciones, obtenidas durante el estudio. Finca Ganadera La Uno S.A, San Carlos, 2017.

Modelo Regresión	n	R^2 ajustado	AIC	BIC	Ajuste	Sesgo	p-valor
1. MR1	785	0,2318	2803,18	2826,51	0,7018	4,63%	<0,0001
2. MR2	785	0,2744	2759,42	2787,41	0,6811	5,06%	<0,0001
3. MR3	785	0,2765	2758,07	2790,73	0,5287	7,61%	<0,0001
4. MR4	785	0,3115	2720,15	2757,47	0,5696	4,41%	<0,0001

Ecuaciones	
1.	$SCCS_{MR1} = 1,6445 + 0,3811CE - 0,0961PL + 0,0203ED$
2.	$SCCS_{MR2} = 0,9629 + 0,5930NPc - 0,0878PL + 0,2376DELc + 0,3560CE$
3.	$SCCS_{MR3} = 1,2954 + 0,4949NPc + 0,1203EDc + 0,2932CE - 0,0785PL + 0,0003CE*DEL$
4.	$SCCS_{MR4} = 1,1580 + 0,5442NPc + 0,0529ED + 0,0018DEL - 0,0153NPc*ED - 0,0885PL + 0,1179CE*NPc$

4.4.2 Algoritmos de *Machine Learning*

Los resultados de los modelos que utilizaron el algoritmo de Naive Bayes se obtuvieron en notación binaria, en los cuales las vacas sanas mostraron el valor de “0”, mientras que las vacas enfermas se identificaron con “1”. El análisis de sensibilidad y especificidad no resultó tan positivo para MLCC y MLCD, como el obtenido por el modelo que utilizó la función de regresión lineal (MLRG), el cual alcanzó los mejores resultados para la predicción de mastitis subclínica con base en niveles de células somáticas. Esto nos sugiere una relación lineal entre los parámetros utilizados y el SCCS. El algoritmo de clasificación de Naive bayes también fue estudiado para la predicción de resultados de inseminación y la tasa de éxito fue baja (Shahinfar *et al.* 2014). Esta deficiencia del modelo para predecir en determinadas condiciones podría estar explicado, ya que este asume que todas las características son independientes, y como se analizó en el ACP, las variables tienen fuerte relación entre ellas (Patil y Sherekar 2013)

Durante el entrenamiento de MLRG se realizaron corridas simultáneas de datos para entrenar el algoritmo, de manera que se obtuvieron los valores de b_0 , m_0 , m_1 , m_2 , m_3 y m_4 . Estos se incluyeron en la ecuación para obtener el resultado de SCCS. EL modelo a prueba se muestra en la siguiente ecuación.

$$SCCS_{MLRG} = 1,1863 + 0,0053*ED + 0,0022*DEL + 0,2616*NP - 0,0810*PL + 0,4102*CE$$

Aunque el análisis de componentes principales indicó que la CE fue de menor peso que el resto de las variables (Figura 9), para el modelo con mayor precisión resultó ser un parámetro de mucho peso, con el valor del coeficiente más alto que asignó el algoritmo. El potencial de esta variable fue evaluado desde 1995 por Nielen y colaboradores, demostrando el potencial que tiene para clasificar vacas sanas y enfermas de mastitis mediante el uso de redes neuronales. Por su parte, DEL y ED fueron variables de mucho valor en el ACP, y aunque forman parte importante en la ecuación, no tuvieron tanto peso en el resultado de SCCS aplicando *machine learning* para la regresión lineal.

Esta tecnología es capaz de encontrar las mejores relaciones debido a que se enfoca en minimizar repetidamente el error, ajustando sus parámetros hasta encontrar la relación donde el error se haga constante mientras tiende a cero.

La prueba de Wilcoxon indica que este fue un modelo ajustado ($p= 0,2761$) y el sesgo resultó en 5,83%, similar a lo obtenido en los modelos alométricos. Sin embargo, la precisión de este modelo superó en al menos 11,5% a todos los modelos.

No se han reportado estudios similares en el ámbito pecuario de Costa Rica y Centroamérica, que mencionen la aplicación de modelos basados en algoritmos de inteligencia artificial para clasificar vacas de acuerdo con el estado sanitario de la ubre. Es por esta razón que el presente estudio abre la ventana para la investigación en este ámbito, de manera que se pueda perfeccionar la implementación del *machine learning* en los sistemas de producción ganaderos para su optimización, con la ventaja de que las RNA son capaces de aprender las relaciones por sí mismo. Y aunque no se pueda mostrar matemáticamente lo que sucede en estas relaciones, los resultados son más acertados a lo que sucede en la realidad.

4.4.3 Prueba de sensibilidad (SEN) y especificidad (ESP)

Cuando se comparó entre los modelos alométricos, los resultados revelaron que los modelos MR1, MR2 y MR3 presentaron una ESP más alta (de 70,0 a 75,0%) en comparación al modelo MR4 (61,3%). Este último destacó, en un paso previo, con los mejores indicadores para estos modelos. En cuanto a SEN, MR4 obtuvo el valor más alto (71,4%) de todos los modelos, lo que indica su buen comportamiento para predecir vacas enfermas en el punto crítico fijado. MLCC obtuvo la SEN más baja (54,5%), y MLRG la ESP más alta (88,8%), esto le da al primero un bajo poder de detección de vacas enfermas y, al segundo el mayor potencial para discriminar vacas sanas. MR1, MR3 y MLCD obtuvieron los mismos resultados en este análisis para la detección de mastitis (Cuadro 6).

Cuadro 6. Análisis de sensibilidad (SEN), especificidad (ESP), precisión y tasa de error, para todos los modelos alométricos y de *machine learning*. Finca Ganadera La Uno S.A, San Carlos, 2017.

Modelo	n	N° variables	SEN (%)	ESP (%)	Precisión (%)	Tasa de error (%)
Modelos de Regresión lineal y linealizados						
1. MR1	157	3	59,7%	75,0%	67,5%	30,3%
2. MR2	157	4	64,9%	70,0%	67,5%	32,4%
3. MR3	157	5	59,7%	75,0%	67,5%	30,3%
4. MR4	157	5	71,4%	61,3%	66,2%	36,0%
Modelos de <i>Machine Learning</i>						
5. MLCD	157	6	59,7%	75,0%	67,5%	30,3%
6. MLCC	157	5	54,5%	77,5%	66,2%	30,0%
7. MLRG	157	5	68,8%	88,8%	79,0%	14,5%

Los resultados sugieren que tanto los modelos de regresión alométricos, como los de ML, que utilizaron el algoritmo de clasificación de Naive Bayes (MLCC y MLCD), no fueron buenos en la predicción de mastitis a partir del Pc determinado, ya que la precisión de estos modelos varió entre 66,2 y 67,5%, mientras tanto la tasa de error fue mayor al 30%. MR4 fue el modelo menos preciso y con mayor tasa de error, a pesar de los parámetros obtenidos entre los modelos alométricos. Por otro lado, se encontró que el modelo con más precisión (79,0%) y menor tasa de error (14,5%) fue MLRG, resultando en valores de SEN y ESP de 68,8 y 88,8%, respectivamente. Otros autores han reportado el uso de redes neuronales para detección de mastitis, utilizando dos variables de CE, dos de rendimiento y los DEL, estas predicciones no fueron exitosas, ya que la tasa de error fue superior al 50% (Cavero *et al.* 2008). Al contrario, en el presente estudio, los modelos elaborados no superaron el 36% de error.

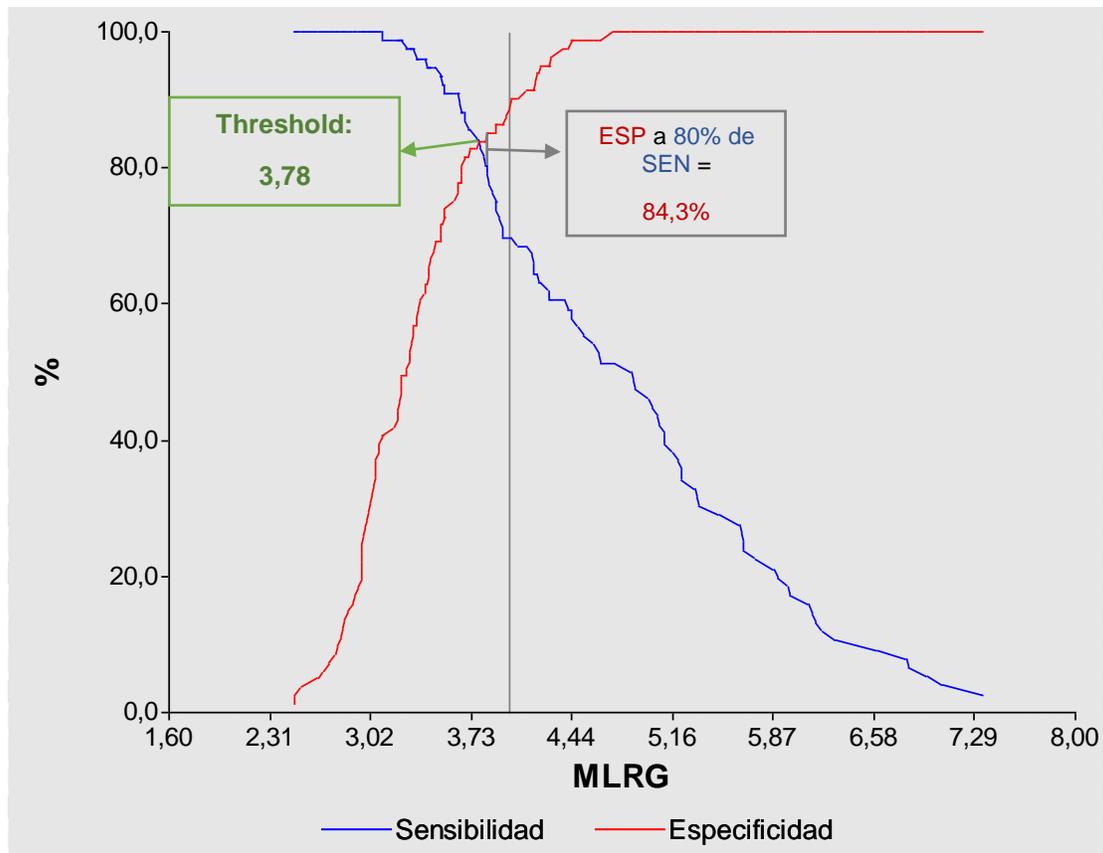


Figura 11. Curvas de sensibilidad (SEN) y especificidad (ESP) del modelo MLRG. Se incluye el *threshold* y se indica la ESP a un 80% de ESP fija. Línea de corte vertical indica el valor de 4 en el eje “x”. Finca Ganadera La Uno S.A, San Carlos, 2017.

En la Figura 11 se muestran las curvas de SEN y ESP para MLRG, mientras que las curvas de los modelos MR se observan en el Anexo 5. Los gráficos incluyen el *threshold* obtenido para cada uno, con el objetivo determinar el valor de la variable pronóstica donde se igualan SEN y ESP. Además, se puede apreciar que el valor de ESP a un 80% de SEN fija fue de 84,3, 52, 56,7, 54,2 y 53% para MLRG, MR1, MR2, MR3 y MR4, respectivamente. Con la utilización de una red neuronal de tipo clasificación, un estudio presenta especificidades de 51,1 y 74,9% a una SEN fijada en 80%, para Pc de 100 000 cs/ml y 400 000 cs/ml, respectivamente. Las variables que se consideraron fueron el rendimiento de leche, la tasa promedio de flujo de leche, el tiempo entre ordeños, la tasa de producción de leche y la CE de cada cuarto (Cavero *et al.* 2008). Ninguno de estos resultados supera la ESP a una SEN fija de 80% de MLRG, con un punto crítico de 3,67 SCCS, equivalente a 159 000 cs/ml.

Por su parte, Ferrero (2014), construyó una red neuronal considerando un mayor set de variables que incluían: productivas (que contemplaban valores absolutos y de variación porcentual con mediciones previas y el promedio del hato), número de lactancias, días en lactación, época de parto, incidencias previas de la enfermedad, dos variables de precipitación, y dos de estrés calórico. La precisión del modelo obtenido fue de 90,74%, y la SEN y ESP fueron de 86,9% y 91,36%, respectivamente. En total, se utilizaron 12 variables y no todas de fácil medición. Mientras tanto, en el presente estudio se obtuvo una precisión aceptable con un set de cinco variables (dos de fácil medición en la línea de leche y tres del registro histórico de los animales).

Utilizando modelos de redes neuronales artificiales y modelos aditivos generalizados, Ankinakatte *et al.* (2013) logró diferenciar entre vacas sanas y enfermas (cultivo bacteriológico). La ESP a un 80% de SEN fija fue cercana a 75%. La ausencia del conteo de células somáticas en el modelo disminuyó la habilidad de predicción en al menos 5%. Ambos algoritmos fueron utilizados para clasificar los datos, siendo estos resultados muy similares a los obtenidos en los modelos de clasificación (MLCC y MLCD). Este autor concluye que los algoritmos de redes neuronales fueron ligeramente mejores para especificidades altas, tendencia que coincide con los resultados obtenidos por los modelos de clasificación (MLC) y regresión (MLRG) (Cuadro 6).

En las curvas ROC se expresa la relación entre la sensibilidad y 1-especificidad, ambas expresadas en porcentaje. El área bajo esta curva constituye el criterio más importante para comparar entre métodos de predicción de una enfermedad, y mientras este valor se aproxima a 1, tiene un mejor valor predictivo, por el contrario, pierde su efectividad cuando llega a 0,5, donde el modelo en estudio no tendría ningún poder de discriminación entre vacas sanas y enfermas. El mejor resultado, obtenido en función del área bajo la curva ROC, fue el modelo MLRG que se muestra en la Figura 12, con un 0,921 del área cubierta. Mientras tanto, los modelos restantes se mantuvieron entre 0,735 y 0,754. En un estudio realizado en vacas Holstein en pastoreo Elizalde *et al.* (2009) donde se desarrolló un modelo de predicción, basado en conductividad eléctrica diferencial (entre cuartos), indica que el área bajo la curva ROC fue de 0,82 y niveles estimados SEN y ESP fueron 78,0 y 79,1%, respectivamente lo cual concuerda con los

resultados obtenidos en esta investigación, donde se encontró evidencia que demuestra que CE es una variable predictora de mastitis. En adición, Shahinfar *et al.* (2014) entrenaron modelos de redes neuronales para determinar el resultado (preñez) de una inseminación artificial en vacas Holstein, estos modelos fueron catalogados por los mismos autores como exitosos, con resultados de 73,6 a 75,6% de área bajo la curva ROC.

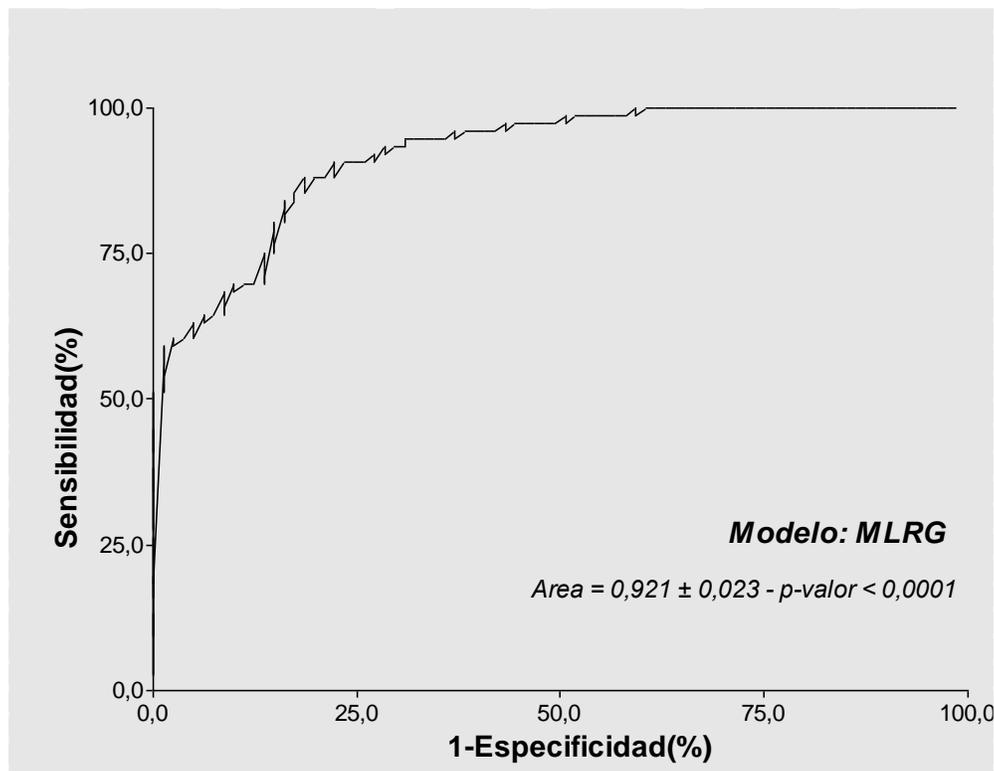


Figura 12. Curva ROC para el modelo MLRG, se incluye el área, error y p-valor de la curva. Se construye con 1-ESP y la SEN, expresados en porcentaje (%). Finca Ganadera La Uno S.A, San Carlos, 2017.

Según Nielen (1995) la utilización de la CE en los modelos de predicción de mastitis para sistemas de detección en línea se ha vuelto muy común desde los inicios de la aplicación de redes neuronales en este campo. Por otra parte, Hassan *et al.* (2009) reportan variaciones a este parámetro, donde se prueba la resistencia eléctrica en lugar de la CE, obteniendo resultados satisfactorios en la identificación de agentes causales de hasta 99% de especificidad en modelos de RNA no supervisados y de 97%-80% en modelos supervisados.

Entender como los algoritmos de inteligencia artificial pueden obtener buenos resultados en la predicción de enfermedades a partir variables sencillas, es todavía una interrogante. Entre las ventajas que pueden ofrecer éstos, sobre los modelos alométricos, radica en que son capaces de aprender de la información que se les provee (datos), buscando constantemente la minimización del error. Tanto las ecuaciones obtenidas en los modelos alométricos como en el MLRG, son ecuaciones lineales o linealizadas. La diferencia radica en que el algoritmo de aprendizaje automático hace reiteradas corridas de datos que le permiten encontrar la intersección con el eje “y” (b_0) y los pesos ($m_0 \dots m_n$) de “n” variables predictoras que se ajusten al mejor resultado de clasificación. Contrario a los modelos alométricos, que buscan, por lo general, maximizar el coeficiente de determinación.

5 CONCLUSIONES

1. Se observó un efecto positivo de los muestreos en la reducción de conteo celular somático, lo que demuestra que el conocimiento de la información del hato por parte del productor aporta en gran medida, en la toma de decisiones, generando mayor beneficio en la producción y precio de la leche en finca.
2. Se determinó que aproximadamente el 50% de las vacas del hato se encontraban enfermas, lo que favoreció el estudio, ya que se contaba con cantidades similares de vacas sanas y enfermas.
3. La disminución en el rendimiento de leche a causa de la mastitis puede estar influenciado por el número de partos, los días en lactación y la edad de la vaca.
4. La conductividad eléctrica demostró poca relación con el score de células somáticas en un análisis previo (ACP), pero cuando se implementó el algoritmo de *machine learning* resultó ser una variable muy importante.
5. Los modelos alométricos fueron muy similares entre ellos, con valores de especificidad a un 80% de sensibilidad fija, cercana a 55%, y áreas de curvas ROC de 75%, aproximadamente.
6. El mejor modelo para la predicción de mastitis a partir de dos variables de rápida medición in situ e indicadores técnicos del animal, fue el algoritmo elaborado con la función de machine learning: "LinearRegression", caracterizado por un área bajo la curva ROC de 0,921 y una especificidad de 84,3 a una sensibilidad fijada de 80%.

6 RECOMENDACIONES

Probar otras variables que mantengan su facilidad de medición (como la temperatura superficial de la ubre) e incluir análisis más robustos para que el acierto en las predicciones sea más elevado.

Identificar cambios repentinos en las dos variables de medición en línea de leche (CE y PL) podría generar mejores predictores de mastitis, así como la inclusión de más variables de rápida medición en los sistemas, tomando en cuenta parámetros de clima y estrés de los animales.

Estudiar el comportamiento de los modelos obtenidos con otros hatos que mantengan niveles de células somáticas menores a los del hato analizado.

Crear una arquitectura de red neuronal que permita explicar mejor el comportamiento de la mastitis en función de estas variables.

Realizar estudios multidisciplinarios que promuevan la integración de investigadores para que, en conjunto, puedan tomar mejores decisiones aportando lo mejor de cada disciplina.

Promover estudio sobre posibles causas generadoras de la alta incidencia de mastitis subclínica en la finca Ganadera La Uno S.A.

7 BIBLIOGRAFÍA

- Addis, MF; Tedde, V; Puggioni, GMG; Pisanu, S; Casula, A; Locatelli, C; Rota, N; Bronzo, V; Moroni, P; Uzzau, S. 2016. Evaluation of milk cathelicidin for detection of bovine mastitis. *Journal of Dairy Science*, 99(10):8250-8258.
- Adriaens, I; Huybrechts, T; Aernouts, B; Geerinckx, K; Piepers, S; De Ketelaere, B; Saeys, W. 2018. Method for short-term prediction of milk yield at the quarter level to improve udder health monitoring. *Journal of dairy science*. 101(11): 10327-10336.
- Alfonso, D; Zanette, J; Ruiz, K; Peña, J; González, Y; Reinoso, M. 2017. Situación de la mastitis subclínica y evaluación de los procesos lecheros en vaquerías de la provincia Villa Clara, Cuba. *Revista de Salud Animal*, 39(3): 00-00.
- Ankinakatte, S; Norberg, E; Løvendahl, P; Edwards, D; & Højsgaard, S. 2013. Predicting mastitis in dairy cows using neural networks and generalized additive models: A comparison. *Computers and Electronics in Agriculture*. 99: 1-6.
- Arturo, AA; Jacinto, BP; Eulogio, PB; Alfonso, AF; Pablo, TM. 2014. Prevalence of subclinical mastitis in municipalities in the region cienega jalisco state. *Abanico Veterinario*. 4(1): 24-31.
- Augspurger, T; Bartak, C; Cloud, P; Hayden, A; Hoyer, S; McKinney, W; Reback, J; Ella, C; Horikoshi, M; Bossche, J. 2019. Pandas. Consultado el 3 de enero del 2019. Obtenido de: <https://pandas.pydata.org/about.html>
- Balas, C. 2009. Review of biomedical optical imaging—a powerful, non-invasive, non-ionizing technology for improving in vivo diagnosis. *Measurement science and technology*, 20(10):104020.
- Batavani, RA; Asri, S; Naebzadeh, H. 2007. The effect of subclinical mastitis on milk composition in dairy cows. *Iranian Journal of Veterinary Research* 8(3):205-211.
- Bytyqi, H; Zaugg, U; Sherifi, K; Hamidi, A; Gjonbalaj, M; Muji, S; Mehmeti, H. 2010. Influence of management and physiological factors on somatic cell count in raw cow milk in Kosova. *Vet Arhiv*, 80(10):173-183.

- Carvajal-Hernández, M; Valencia-Heredia, ER; Segura-Correa, JC. 2002. Duración de la lactancia y producción de leche de vacas Holstein en el Estado de Yucatán, México. *Revista Biomédica*. 13(1): 25-31.
- Cavero, D. 2007. Automated mastitis detection in dairy cows using different statistical methods. Tesis Ph.D. Kiel Alemania. Universidad de Kiel. 89p.
- Cavero, D; Tölle, KH; Henze, C; Buxadé, C; Krieter, J. 2008. Mastitis detection in dairy cows by application of neural networks. *Livestock Science*. 114(2-3): 280-286.
- COOPROLE (Cooperativa de Productores de Leche Dos Pinos R.L.). 2016. Reglamento de Recibo de Leche. 26 p.
- De Mol, RM; Ouweltjes, W. 2001. Detection model for mastitis in cows milked in an automatic milking system. *Preventive veterinary medicine*. 49(1-2): 71-82.
- Di Rienzo J.A; Casanoves, F; Balzarini, M.G; Gonzalez, L; Tablada, M; Robledo, C.W. 2017. Infostat (en línea). Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina. Disponible en <http://www.infostat.com.ar>
- Digiovani, DB; Borges, MHF; Galdioli, VHG; Matias, BF; Bernardo, GM; da Rocha Silva, T; Cruz, P; Barca, FA; Guiselli, F; Koetz, C; de Azambuja Ribeiro, EL. 2016. Infrared thermography as diagnostic tool for bovine subclinical mastitis detection. *Revista Brasileira de Higiene e Sanidade Animal: RBHSA*, 10(4):685-692.
- Elizalde, EF; Signorini, ML; Canavesio, VR; Cuatrin, A; Tarabla, HD; Calvino, LF. 2009. Medición de la Conductividad Eléctrica en Leche Como Método Diagnóstico de Mastitis Subclínica Bovina. *Revista FAVE - Ciencias Veterinarias* 8(1):5-28.
- Farmonline Weather (Australia). 2017. Station Details: Kyabram northern Country, Victoria (en línea, sitio web). Consultado 15 mayo. 2017. Disponible en <http://www.farmonlineweather.com.au/climate/station.jsp?lt=site&lc=80091>
- Fernández, ML; Ramírez, JP; Chaves, C; Arias, ML. 2008. Disminución en la incidencia de mastitis en ganado vacuno con la aplicación de un sellador de barrera experimental. *Agronomía Costarricense: Revista de ciencias agrícolas*, 32(1):107-112.

- Ferrero, M. 2014. 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. Paper presented at the XLIII Jornadas Argentinas De Informática E Investigación Operativa (43JAIIO)-XV Argentine Symposium on Technology (AST) (Buenos Aires).
- García, CN. 1999. Exploración del uso potencial de las redes neuronales para la clasificación de la leche según rango celular. Tesis de grado presentada como parte de los requisitos para optar al Grado de licenciado en medicina veterinaria. Universidad Austral de Chile. 48p.
- Gonçalves, JL; Cue, RI; Botaro, BG; Horst, JA; Valloto, AA; Santos, MV. 2018a. Milk losses associated with somatic cell counts by parity and stage of lactation. *Journal of dairy science*. 101(5): 4357-4366.
- Gonçalves, JL; Kamphuis, C; Martins, CMMR; Barreiro, JR; Tomazi, T; Gameiro, AH; dos Santos, MV. 2018b. Bovine subclinical mastitis reduces milk yield and economic return. *Livestock Science*. 210: 25-32.
- Gráf, M; Mikó, E. 2015. Analysis of Mastitis in Holstein- Friesian Cows and Economic Effects of Mastitis. *Agricultural Management / Lucrari Stiintifice Seria I, Management Agricol* 17(1):294-297.
- Green, LE; Schukken, YH; Green, MJ. 2006. On distinguishing cause and consequence: Do high somatic cell counts lead to lower milk yield or does high milk yield lead to lower somatic cell count?. *Preventive veterinary medicine*. 76(1-2): 74-89.
- Guerrero-Quiceno, J; Campos-Gaona, R; Vélez-terranova, M. 2018. Factores ambientales asociados y métodos de determinación de células somáticas en leche de tanque. *Biotecnología en el Sector Agropecuario y Agroindustrial*, 16(1):80-90.
- Guimarães, JL; Brito, MA; Lange, CC; Silva, MR; Ribeiro, JB; Mendonça, LC; Mendonça, JMF; Souza, GN. 2017. Estimate of the economic impact of mastitis : a case study in a Holstein dairy herd under tropical conditions. *Preventive Veterinary Medicine*.

- Wickham, H. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
- Hassan, KJ; Samarasinghe, S; Lopez-Benavides, MG. 2009. Use of neural networks to detect minor and major pathogens that cause bovine mastitis. *Journal of Dairy Science*, 92(4), 1493-1499.
- Hertl, JA; Schukken, YH; Tauer, LW; Welcome, FL; Gröhn, YT. 2018. Does clinical mastitis in the first 100 days of lactation predict increased mastitis occurrence and shorter herd life in dairy cows?. *Journal of dairy science*, 101(3): 2309-2323.
- Hoernig, KJ; Donovan, DM; Pithua, P; Williams, F; Middleton, JR. 2016. Evaluation of a lysostaphin-fusion protein as a dry-cow therapy for *Staphylococcus aureus* mastitis in dairy cattle. *Journal of dairy science*, 99(6):4638-4646.
- Huijps, K; Lam, TJ; Hogeveen, H. 2008. Costs of mastitis: facts and perception. *The Journal of dairy research*, 75(1):113-120.
- Kamphuis, C. 2010. Making sense of sensor data: Detecting clinical mastitis in automatic milking systems. Dissertation Faculty of Veterinary Medicine. 11-21.
- Kamphuis, C; Rue, BD; Eastwood, CR. 2016. Field validation of protocols developed to evaluate in-line mastitis detection systems. *Journal of dairy science*, 99(2):1619-1631.
- Kamphuis, C; Sherlock, R; Jago, J; Mein, G; Hogeveen, H. 2008. Automatic detection of clinical mastitis is improved by in-line monitoring of somatic cell count. *Journal of Dairy Science*, 91(12):4560-4570.
- Kaşıkçı, G; Çetin, Ö; Bingöl, EB; Gündüz, MC. 2012. Relations between electrical conductivity, somatic cell count, California mastitis test and some quality parameters in the diagnosis of subclinical mastitis in dairy cows. *Turkish Journal of Veterinary and Animal Sciences*, 36(1):49-55.
- Koskinen, MT; Holopainen, J; Pyörälä, S; Bredbacka, P; Pitkälä, A; Barkema, HW; Bexiga, R; Roberson, J; Sølverød, L; Piccinini, R; Kelton, D; Lehmusto, H; Niskala, S; Salmikivi, L. 2009. Analytical specificity and sensitivity of a real-time

- polymerase chain reaction assay for identification of bovine mastitis pathogens. *Journal of Dairy Science*, 92(3):952-959.
- Lavon, Y; Kaim, M; Leitner, G; Biran, D; Ezra, E; Wolfenson, D. 2016. Two approaches to improve fertility of subclinical mastitic dairy cows. *Journal of dairy science*, 99(3):2268-2275.
- Lavon, Y; Leitner, G; Voet, H; Wolfenson, D. 2010. Naturally occurring mastitis effects on timing of ovulation, steroid and gonadotrophic hormone concentrations, and follicular and luteal growth in cows. *Journal of dairy science*, 93(3):911-921.
- Lee, KH; Lee, JW; Wang, SW; Liu, LY; Lee, MF; Chuang, ST; Chi, CH. 2008. Development of a novel biochip for rapid multiplex detection of seven mastitis-causing pathogens in bovine milk samples. *Journal of veterinary diagnostic investigation*, 20(4):463-471.
- López, R. F., & Fernández, J. M. F. 2008. Las redes neuronales artificiales. Netbiblo. 152p.
- Mollenhorst, H; Van der Tol, PPJ; Hogeveen, H. 2010. Somatic cell count assessment at the quarter or cow milking level. *Journal of dairy science*, 93(7):3358-3364.
- Momena, K; Cameron, EFC; Lyons AN; Thomson, PC; Kendra, LK; García SC. 2017. Early detection of clinical mastitis from electrical conductivity data in an automatic milking system. *Animal Production Science*, 57:1226-1232.
- Mora, MG; Vargas, B; Romero, JJ; Camacho, J. 2015. Factores de riesgo para la incidencia de MC en ganado lechero de Costa Rica. *Agronomía Costarricense*, 39(2):77-89.
- Nääs, IA; Garcia, RG; Caldara, FR. 2014. Infrared thermal image for assessing animal health and welfare. *JABB-Online Submission System*, 2(3):66-72.
- Nicholas, RA; Fox, LK; Lysnyansky, I. 2016. *Mycoplasma mastitis* in cattle: To cull or not to cull. *The Veterinary Journal*, 216:142-147.

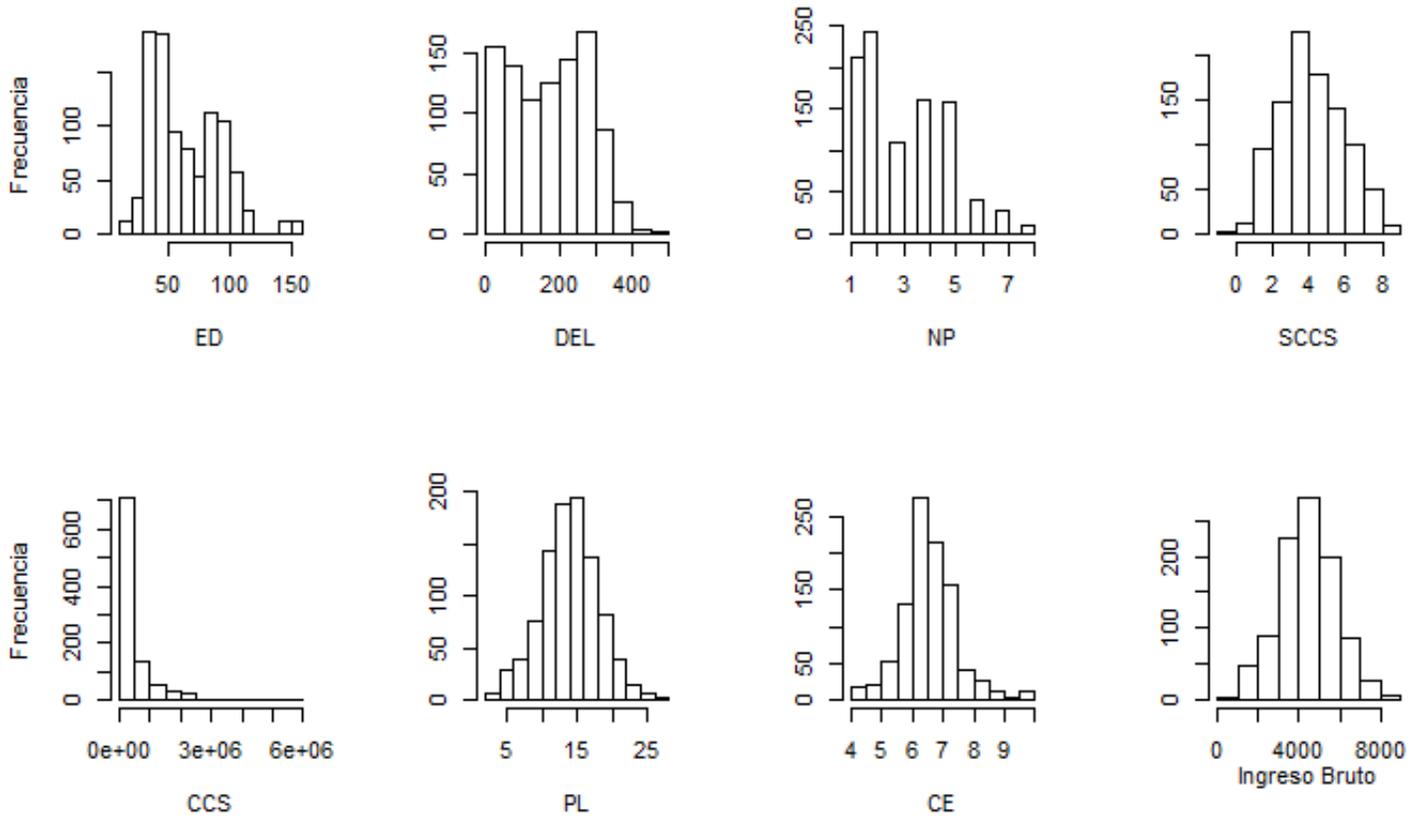
- Nielen, M; Spigt, MH; Schukken, YH; Deluyker, HA; Maatje, K; Brand, A. 1995. Application of a neural network to analyse on-line milking parlour data for the detection of clinical mastitis in dairy cows. doi://doi.org/10.1016/0167-5877(94):00405-8
- Nielsen, C. 2009. Economic impact of mastitis in dairy cows. Tesis Ph.D. Uppsala, Swedish. University of Agricultural Sciences of Uppsala.
- Norberg, E; Rogers, GW; Goodling, C; Cooper, JB; Madsen, P. 2004. Genetic Parameters for Test-Day Electrical Conductivity of Milk for First-Lactation Cows from Random Regression Models. *J. Dairy Sci.* 87:1917–1924.
- Oviedo-Boyso, J; Valdez-Alarcón, JJ; Cajero-Juárez, M; Ochoa-Zarzosa, A; López-Meza, JE; Bravo-Patiño, A; Baizabal-Aguirre, VM. 2007. Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. *Journal of Infection*, 54(4):399-409.
- Patil, TR; Sherekar, SS. 2013. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. *International journal of computer science and applications.* 6(2): 256-261.
- Pedraza, C. 1991. Efecto de la mastitis clínica sobre la producción de leche. *Agricultura Técnica.* 51:298-305.
- Pedregosa, F; Varoquaux, G; Gramfort, A; Michel, V; Thirion, B; Grisel, O; Blondel, M; Prettenhofer, P; Weiss, R; Dubourg, V; Vanderplas, J; Passos, A; Cournapeau, D; Brucher, M; Perrot, M; Duchesnay, E. 2011. Scikit-learn: Machine Learning in Python. *JMLR.* 12: 2825-2830.
- Polat, B; Colak, A; Cengiz, M; Yanmaz, LE; Oral, H; Bastan, A; Kaya, S; Hayirli, A. 2010. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. *Journal of Dairy Science*, 93(8):3525-3532.
- Porporatto, C; Felipe, V. 2010. Mastitis, confort animal y calidad de leche. Obtenido de: <https://ebookcentral.proquest.com>

- R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Rees, A; Fischer-Tenhagen, C; Heuwieser, W. 2017. Udder firmness as a possible indicator for clinical mastitis. *Journal of Dairy Science*, 100(3):2170-2183.
- Ribeiro, ADC; dos Santos, JS; Zanol, D; Lombarde, LNL; Bruzaroski, SR; Ludovico, A; de Santana, EHW. 2016. Evaluation of an electrical conductivity portable device as an alternative for subclinical mastitis detection. *Revista de Salud Animal* 38(2):31-135.
- Roth, Z; Wolfenson, D. 2016. Comparing the effects of heat stress and mastitis on ovarian function in lactating cows: basic and applied aspects. *Domestic Animal Endocrinology*, 56: 218-227.
- Rowbotham, RF; Ruegg, PL. 2016. Associations of selected bedding types with incidence rates of subclinical and clinical mastitis in primiparous Holstein dairy cows. *Journal of dairy science*, 99(6):4707-4717.
- Shahinfar, S; Page, D. Guenther, J; Cabrera, V; Fricke, F; Weigel, K. 2014. Prediction of insemination outcomes in Holstein dairy cattle using alternative machine learning algorithms. *J. Dairy Sci.* 97 :731–742.
- Sharma, N; Kang, TY; Lee, SJ; Kim, JN; Hur, CH; Ha, JC; Jeong, DK. 2013. Status of bovine mastitis and associated risk factors in subtropical Jeju Island, South Korea. *Tropical animal health and production*. 45(8): 1829-1832.
- Sharma, N; Pandey, V; Sudhan, NA. 2010. Comparison of Some Indirect Screening Tests for Detection of Subclinical Mastitis in Dairy Cows. *Bulgarian Journal of Veterinary Medicine*. 13(2):98-103.
- Sharma, N; Singh, K; Bhadwal MS. 2011. Relationship of Somatic Cell Count and Mastitis: An Overview. *Asian-Aust. J. Anim. Sci* 24(3):429-438.

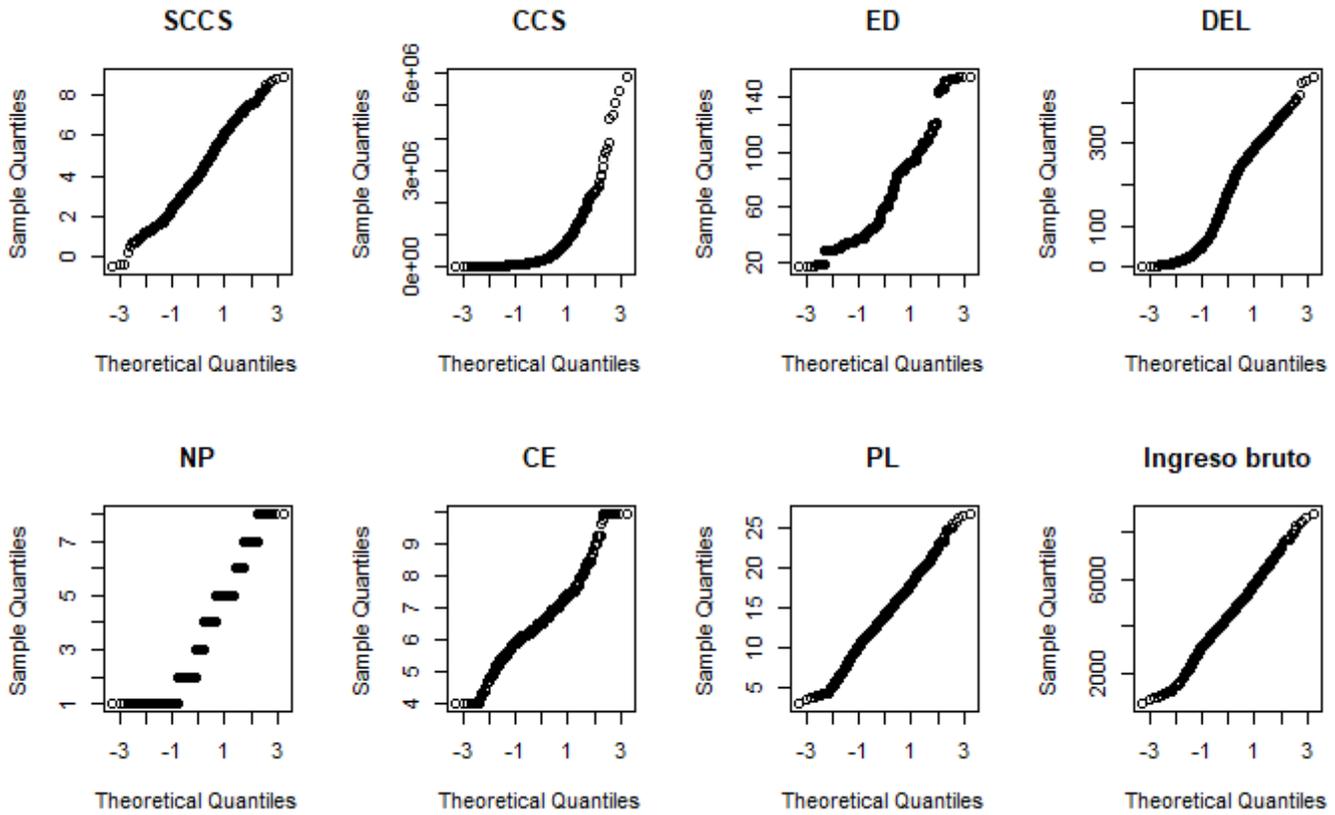
- Shinea, P; Murphya, MD; Uptonc, J; Scullyb, T. 2018. Machine-learning algorithms for predicting on-farm direct water and electricity consumption on pasture based dairy farms. *Computers and Electronics in Agriculture* 150 (2018) 74-87.
- Shook, GE; Schutz, MM. 1994. Selection on somatic cell score to improve resistance to mastitis in the United States. *Journal of Dairy Science*. 77(2): 648-658.
- Skene, JKM; Harford, LB. 1964. Soils and land use in the Rochester and Echuca Districts, Victoria. Technical Bulletin No17. Department of Agriculture, Victoria.
- Smith, K.L., Hogan, J.S. 2001. The world of mastitis. Proceeding of the 2nd International Symposium on mastitis and milk quality, Vancouver, Canada pp 1 – 12.
- Vásquez, JA; Novoa, CF; Carulla, JE. 2014. Efecto del recuento de células somáticas sobre la aptitud quesera de la leche y la calidad fisicoquímica y sensorial del queso campesino. *Revista de la Facultad de Medicina Veterinaria y de Zootecnia*. 61(2):171-185.
- Viguiet, C; Arora, S; Gilmartin, N; Welbeck, K; O’Kennedy, R. 2009. Mastitis detection: current trends and future perspectives. *Trends in biotechnology*, 27(8):486-493.
- Waterman, C; Nelson, N. 2016. Dairy farm monitor report, annual report (2015-16). Victorian Government Department of Economic Development, Jobs, Transport and Resources. Melbourne, Victoria, Australia. 100p.
- Yang, Y; Cao, S; Zhao, X; Huang, D; Zhao, H; Cheng, G. 2016. Determination of changes in bovine plasma and milk proteins during naturally occurring *Escherichia coli* mastitis by comparative proteomic analysis. *Animal Production Science*. 56():1888-1896.
- Zambrano, WJ; de Pinho JA. 2008. Evaluación de la glándula mamaria y composición química de la leche en vacas primíparas mestizas lecheras en el parto, hasta el quinto mes de la lactación. *Revista Científica*, 18(5): 562-569.
- Zhao, X; Lacasse, P. 2014. Mammary tissue damage during bovine mastitis: Causes and control 1. *Journal of Animal Science* 86(13):57-65.

8 ANEXOS

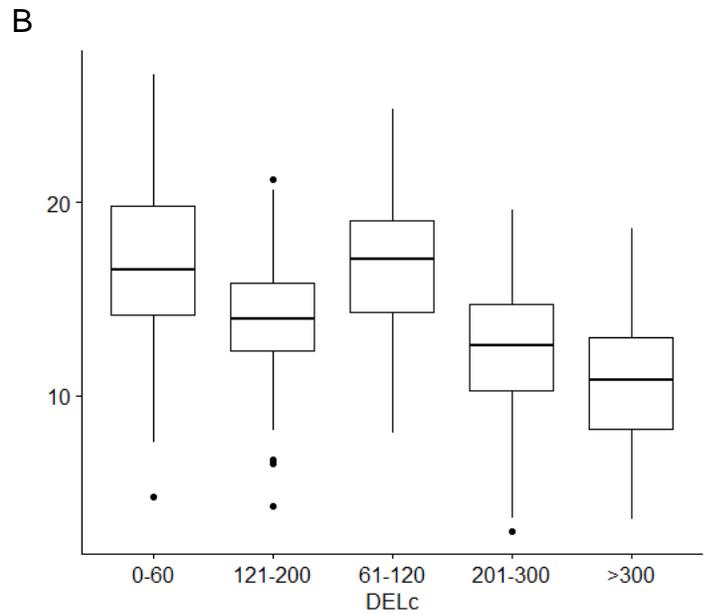
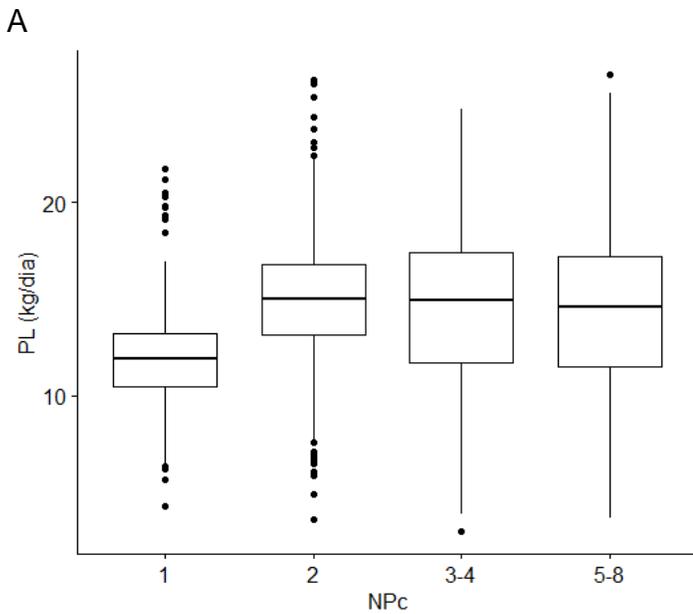
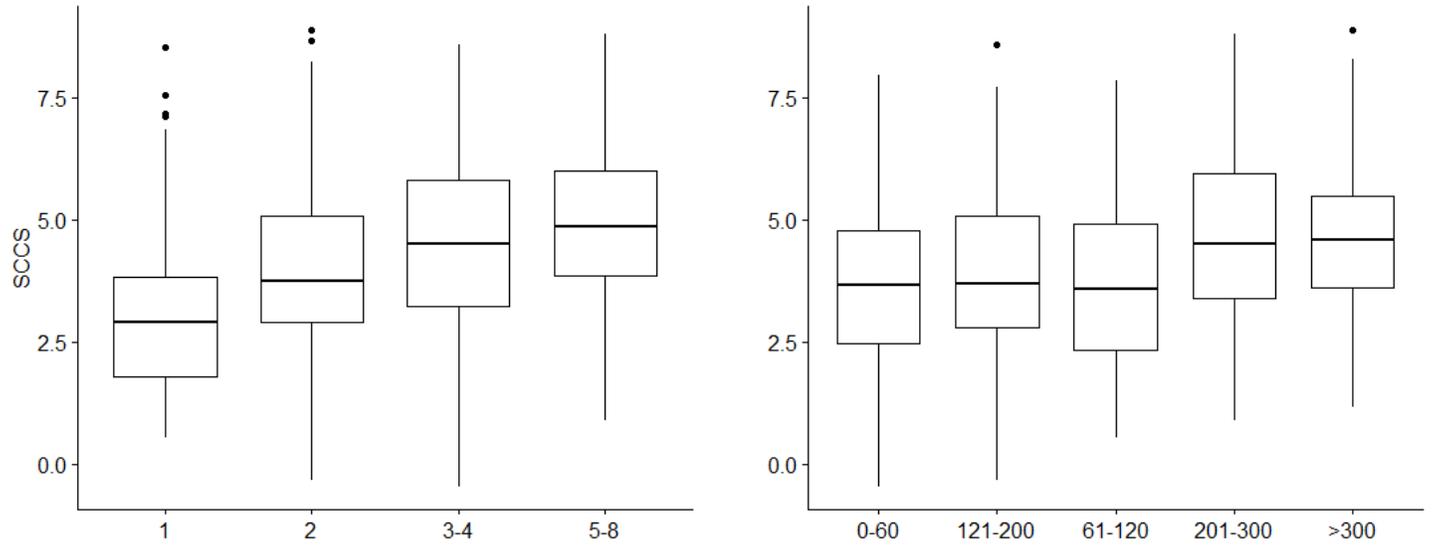
Anexo 1. Histogramas de frecuencia para las variables utilizadas en los modelos de predicción.



Anexo 2. QQ-Plots de las variables utilizadas en los modelos de predicción. SCCS, CE y PL no tienen una desviación observable de distribución normal.



Anexo 3. Boxplots* de SCCS y Producción de Leche (PL) vrs número de parto categorizado (NPc) y días en lactación categorizado (DELc). A) Comportamiento de SCCS respecto al número de partos categorizados. B) Comportamiento del SCCS respecto a los días en lactación. C) Producción de leche respecto al número de partos categorizado. D) Producción de leche respecto a los días en lactación.



C

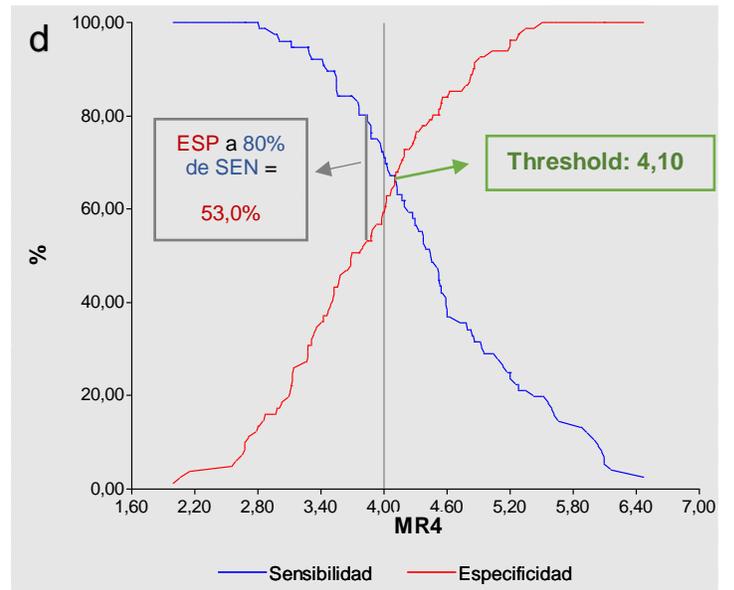
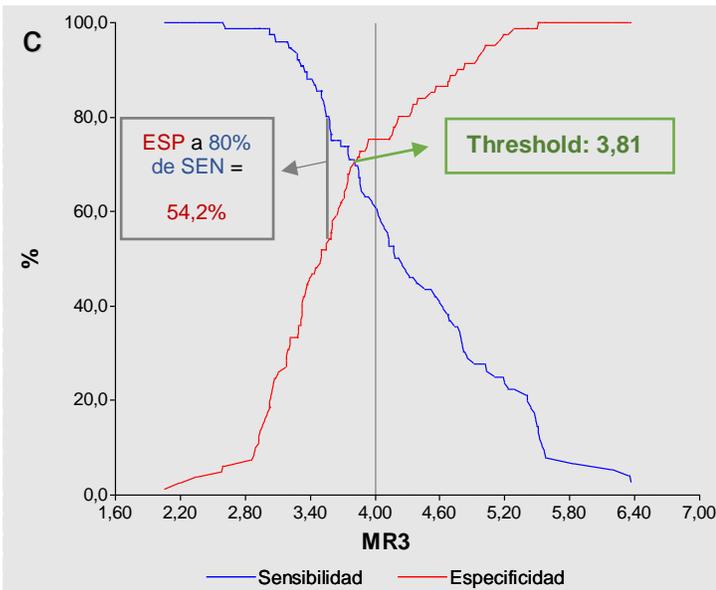
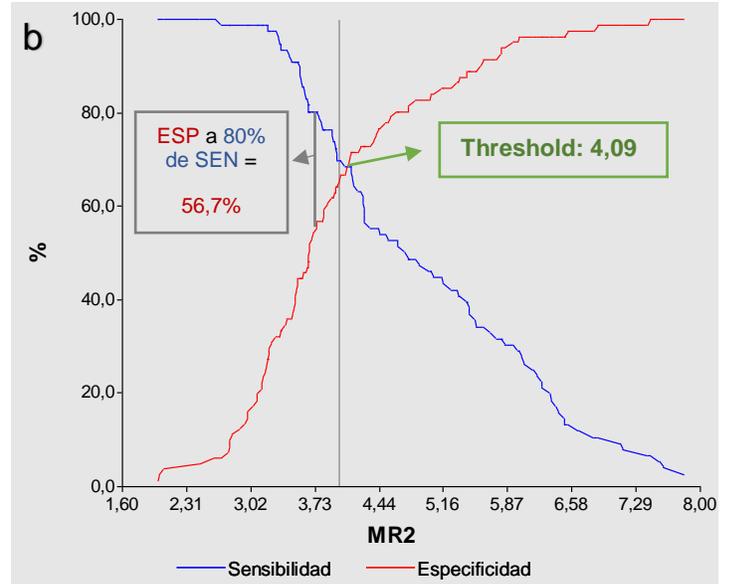
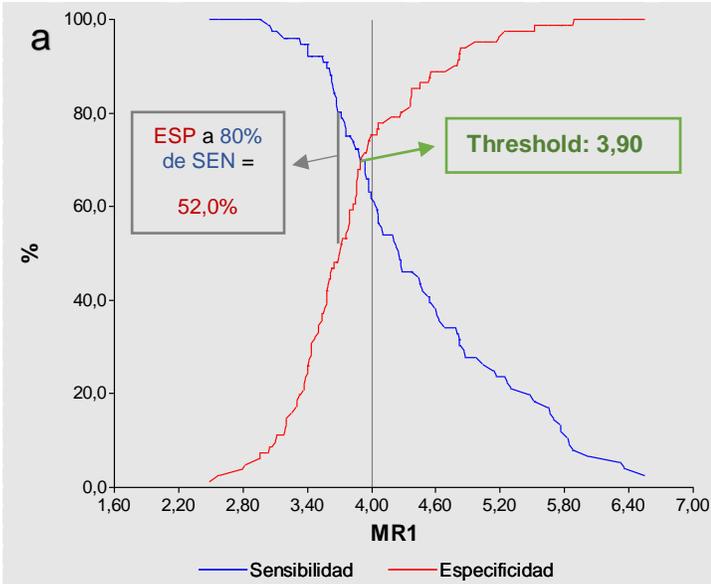
D

*en un boxplot, la línea dentro de la caja representa la mediana, la línea inferior de la caja representa el primer cuartil, la línea superior de la caja representa el tercer cuartil y los bigotes (whiskers) representan 1.5 veces el rango intercuartil + el tercer cuartil y 1.5 veces el rango intercuartil – el primer cuartil para el bigote superior e inferior respectivamente. Visualmente los boxplots funcionan para evaluar el supuesto de igualdad de varianzas por grupos en las variables.

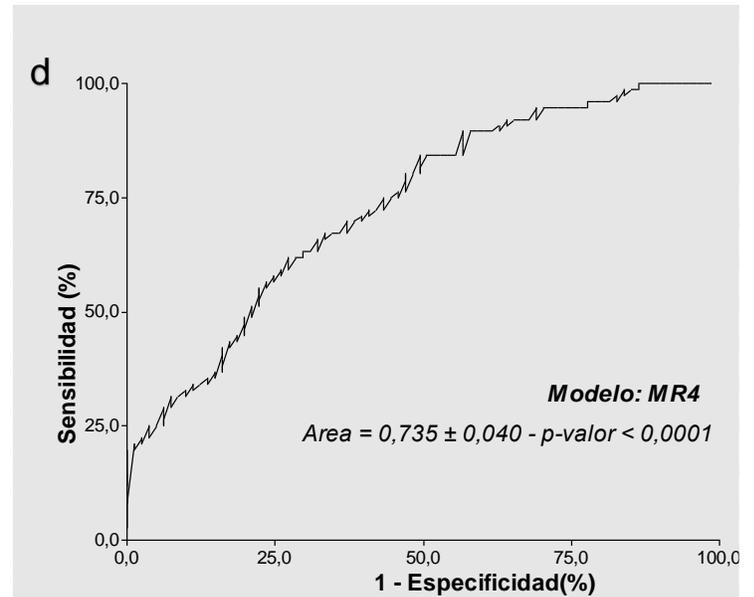
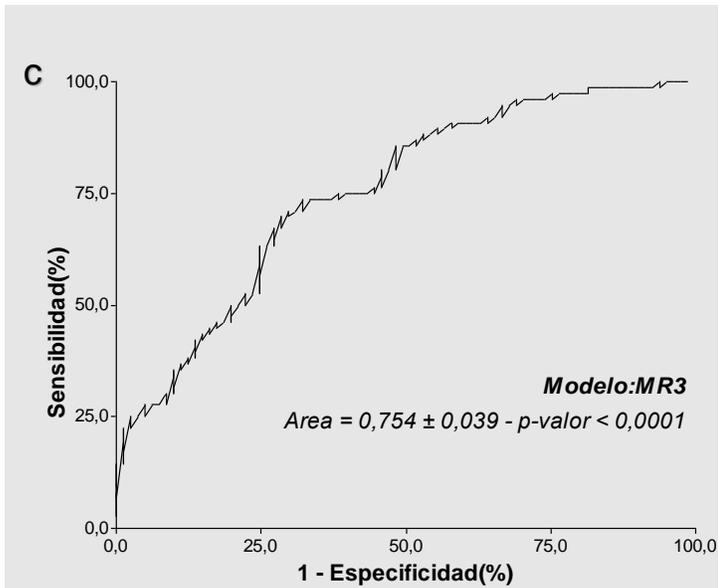
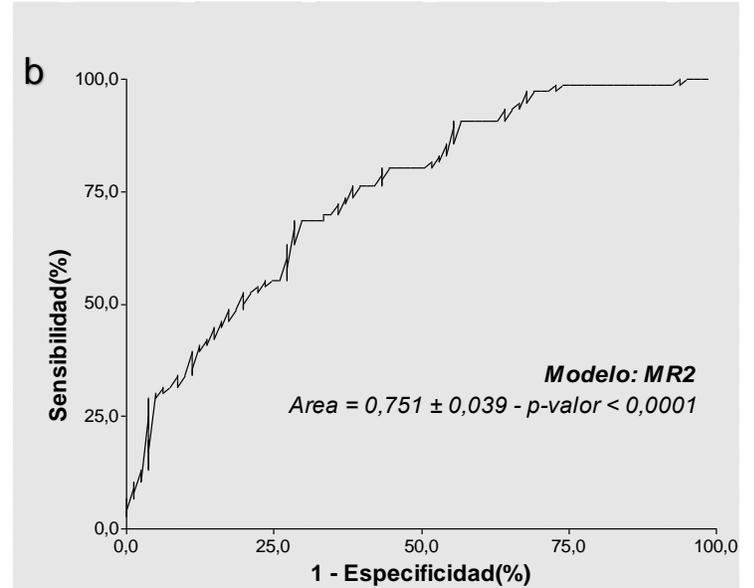
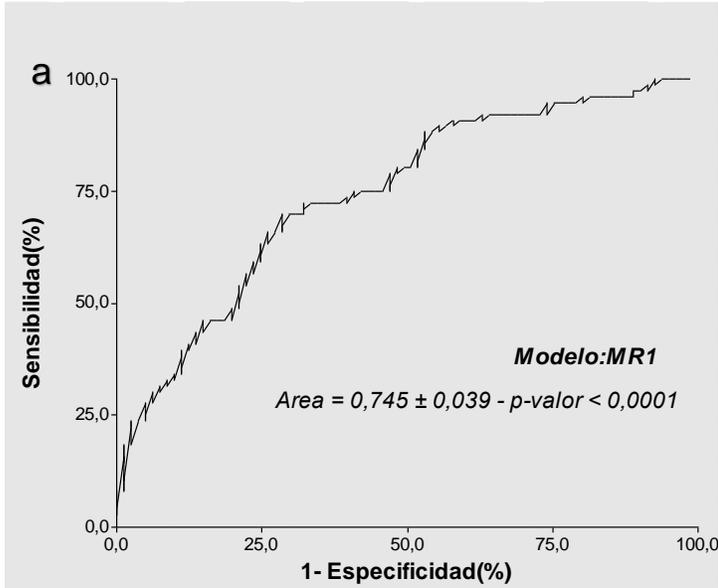
Anexo 4. Afectación porcentual del precio de la leche en función del conteo celular somático (COOPROLE, 2016).

CCS (cs/ml)		Bonificación/ Castigo
De	hasta	
0	100 000	4,00%
100 001	125 000	3,75%
125 001	150 000	3,50%
150 001	175 000	3,25%
175 001	200 000	3,00%
200 001	225 000	2,50%
225 001	250 000	2,00%
250 001	275 000	1,80%
275 001	300 000	1,60%
300 001	325 000	1,40%
325 001	350 000	1,20%
350 001	375 000	1,00%
375 001	400 000	0,80%
400 001	425 000	0,60%
425 001	450 000	0,40%
450 001	475 000	0,20%
475 001	500 000	0,00%
500 001	525 000	-0,50%
525 001	550 000	-1,00%
550 001	575 000	-1,50%
575 001	600 000	-2,00%
600 001	625 000	-2,50%
625 001	650 000	-3,00%
650 001	675 000	-3,50%
675 001	700 000	-4,00%
700 001	1 000 000	-10,00%
1 000 001	9 900 000	-25,00%

Anexo 5. Curvas de sensibilidad (SEN) y especificidad (ESP) del modelo MLRG. Se incluye el threshold y se indica la ESP a un 80% de ESP fija. Línea de corte vertical indica el valor de 4 en el eje "x". Se muestran los modelos: a) MR1, b) MR2, c) MR3 y d) MR4.



Anexo 6. Curvas ROC de los modelos alométricos MR1 (a), MR2 (b), MR3 (c), MR4 (d).
Se incluye el área, error y p-valor de la curva.



Anexo 7. Programación para el entrenamiento del modelo de clasificación de Naive bayes.

```
In [ ]: def training_model(array_to_train):
df = pd.read_csv("./mastitis.csv") # cargar los datos sobre mastitis
df = df.set_index('ID_muestra')
feature_col_names = ['ED',
                    'DEL',
                    'NP',
                    'PL',
                    'CE',
                    ]
predicted_class_name = ['Resultado']
X = (df[feature_col_names].values)
y = (df[predicted_class_name].values)
split_test_size = 0.3
X_test, y_test = train_test_split(X, y, test_size=split_test_size, random_state=45)
nb_model = GaussianNB()
nb_model.fit(X, y.ravel())
nb_predict_test = nb_model.predict(X_test)
confusion_matrix = (metrics.confusion_matrix(y_test, nb_predict_test))
classification_report = (metrics.accuracy_score(y_test, nb_predict_test))
print(classification_report)
nb_predict_test_entered = nb_model.predict([array_to_train])
prediction_report = [
    {"test_result": nb_predict_test_entered},
    {"class_report": classification_report},
    {"confusion_matrix": confusion_matrix}
]
return prediction_report // luego se usa esta funcion para llamarla desde el model predict, vinculado a googlesheet
s
```

Anexo 8. Programación para el entrenamiento de la función LinearRegression.

```
In [1]: import pandas as pd
import numpy as np
from sklearn import linear_model
```

```
In [2]: df = pd.read_csv("mastitis2.csv")
```

```
In [3]: # crear el objeto de regresion lineal
reg = linear_model.LinearRegression()
#crear variables independientes y variable de respuesta
reg.fit(df[['ED', 'DEL', 'NP', 'PL', 'CE']], df.SCCS)
```

```
Out[3]: LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None,
normalize=False)
```

```
In [4]: #ecuacion
#SCCS = 0.00526885*ED + 0.00223176*DEL + 0.26163467*NP + -0.08995589* PL + 0.41015837* CE + 1.1863627017735534
#imprimir coeficientes m1, m2, m3
# m1 = 6.2, m2 = 2.06...
# estos valores se ven en el array que se imprime a continuacion
reg.coef_
```

```
Out[4]: array([ 0.00526885,  0.00223176,  0.26163467, -0.08995589,  0.41015837])
```

```
In [5]: #sccs = 6.28145154e-03*ED + 2.06807602e-03*DEL + 1.19539235e-01*NP + -2.24793187e-02*PL + 1.35827066e-01*CE + 1.77858945e-06
+CSS +1.6223367541315907
# ahora vamos a imprimir el valor de B
# b = 1.62234
reg.intercept_
```

```
Out[5]: 1.1863627017735534
```

```
In [6]: # aca se meten en orden los valores de ED, DEL, NP, PL, CE, CSS RESPESTIVAMENTE
#Y se predice el valor de sccs
# en este caso dio 3.91 el valor de css
reg.predict([[92, 49, 5, 13, 5]]) #5.01
```

```
Out[6]: array([3.96999136])
```

```
In [7]: from sklearn import metrics
# definir valores predcidos y valores reales
y_true = [4.83, 5.01, 5.00]
y_pred = [3.99720737, 4.01933303, 3.8369086]

# calcular MAE, MSE, RMSE
print(metrics.mean_absolute_error(y_true, y_pred))
print(metrics.mean_squared_error(y_true, y_pred))
print(np.sqrt(metrics.mean_squared_error(y_true, y_pred)))
print(((4.83-3.99720737)/4.83)*100) #83% efectividad, 17% error
print(((5.01-4.01933303)/5.01)*100) #81% efectividad, 19% error
```

```
0.9955169999999999
1.0092487382617523
1.0046137258975474
17.242083436853004
19.773791816367257
```