

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
Programa de Pós-Graduação em Zootecnia

Gabriel Machado Dallago

**AVALIAÇÃO DE MODELOS PARA PREDIZER A PRODUÇÃO DE
LEITE NO PRIMEIRO CONTROLE LEITEIRO DE PRIMÍPARAS**

Diamantina
2018

Gabriel Machado Dallago

**AVALIAÇÃO DE MODELOS PARA PREDIZER A PRODUÇÃO DE LEITE NO
PRIMEIRO CONTROLE LEITEIRO DE PRIMÍPARAS**

Dissertação apresentada ao Programa de Pós-Graduação em Zootecnia da Universidade Federal dos Vales do Jequitinhonha e Mucuri, como requisito parcial para a obtenção do título de Mestre.

Orientadora: Profa. Dra. Darcilene Maria de Figueiredo

**DIAMANTINA - MG
2018**

Elaborado com os dados fornecidos pelo(a) autor(a).

D144a

Dallago, Gabriel Machado

Avaliação de modelos para predizer a produção de leite no primeiro controle leiteiro de primíparas / Gabriel Machado Dallago . – Diamantina, 2018.

58 p. : il.

Orientadora: Darcilene Maria de Figueiredo

Coorientadores: Paulo César de Resende Andrade e Roseli Aparecida dos Santos Summary

Dissertação (Mestrado – Curso de Pós-Graduação em Zootecnia) - Universidade Federal dos Vales do Jequitinhonha e Mucuri.

1. Aprendizado de máquina. 2. Distúrbios metabólicos. 3. Novilhas. 4. Zootecnia de precisão. I. Figueiredo, Darcilene Maria de. II. Andrade, Paulo César de Resende. III. Summary, Aparecida dos Santos. IV. Título. V. Universidade Federal dos Vales do Jequitinhonha e Mucuri.

CDD 636.21

GABRIEL MACHADO DALLAGO

**Avaliação de modelos para prever a produção de leite no primeiro
controle leiteiro de primíparas**

Dissertação apresentada ao
MESTRADO EM ZOOTECNIA, nível
de MESTRADO como parte dos
requisitos para obtenção do título de
MAGISTER SCIENTIAE EM
ZOOTECNIA

Orientador (a): Prof.^a Dr.^a Darcilene
Maria De Figueiredo

Data da aprovação : 20/08/2018

Prof.Dr.^a DARCILENE MARIA DE FIGUEIREDO - UFVJM

Prof.Dr.^a ROSELI APARECIDA DOS SANTOS - UFVJM

Prof.Dr. LUCAS LIMA VERARDO - UFVJM

Prof.Dr. LEONARDO SIQUEIRA GLORIA - UENF

Prof.Dr. PAULO CESAR DE RESENDE ANDRADE - UFVJM

GABRIEL MACHADO DALLAGO

**Avaliação de modelos para prever a produção de leite no primeiro
controle leiteiro de primíparas**

Dissertação apresentada ao
MESTRADO EM ZOOTECNIA, nível
de MESTRADO como parte dos
requisitos para obtenção do título de
MAGISTER SCIENTIAE EM
ZOOTECNIA

Orientador (a): Prof.^a Dr.^a Darcilene
Maria De Figueiredo

Data da aprovação : 20/08/2018

DIAMANTINA

AGRADECIMENTOS

Sou hoje o reflexo das incontáveis interações e experiências vividas. Dias se passaram e transformaram-se em semanas, meses e anos, trazendo consigo um enovelado de memórias e crescimento pessoal. Muitos foram os protagonistas ao meu lado nessa jornada, desempenhando papéis nem sempre bem definidos, mas que certamente influenciaram no que sou hoje. Não é tarefa fácil expressar em tão poucas palavras a minha gratidão, essa é então uma humilde tentativa.

Sou grato aos meus pais, Celso Dallago e Natália Machado de Almeida, e ao meu irmão Bruno Machado Dallago pelo apoio e suporte incondicional.

À minha orientadora, Professora Doutora Darcilene Maria de Figueiredo, serei sempre grato pela leveza, liberdade, coragem, confiança e incentivo ao longo desses 24 meses. Não poderia ter sido diferente.

Sou imensuravelmente grato ao Professor Doutor Paulo César de Resende Andrade e à Professora Doutora Roseli Aparecida dos Santos pelos sábios conselhos, energia e incansável apoio, não medindo esforços em ajudar.

Sou grato pela oportunidade de ter sido membro do Nepel (Núcleo de Estudos em Pecuária Leiteira) e do Ruminar (Núcleo de Estudos e Difusão do Conhecimento em Nutrição de Ruminantes). Foi sem dúvida alguma onde tive risadas, viagens, conversas e experiências inesquecíveis!

À Elizângela Aparecida Saraiva! O que seria de mim sem a sua ajuda sempre acompanhada de um bom humor invejável!

À CAPES/CNPq pelo apoio financeiro durante o mestrado.

À Valacta pela disponibilização do conjunto de dados utilizado nesse trabalho.

Ao meu *cluster* Priscila Junia Rodrigues da Cruz e Hellen Cristina Ferreira Antunes! Meses de convivência amigável marcados por festas de aniversário não convencionais, acontecimentos inesperados, surpresas positivas e outras nem tanto, mas sempre a certeza de ter com quem contar nos momentos Berola's e nos momentos não tão Berola's assim.

Deixo aqui o meu muito obrigado a todos vocês!

RESUMO

O período de transição é marcado por profundas alterações morfofisiológicas no corpo do gado leiteiro. A maioria dos distúrbios metabólicos e doenças infecciosas são diagnosticados durante o parto, comprometendo o sucesso de toda a lactação. Formas de se avaliar o manejo de transição de multíparas têm sido propostas, muito embora o mesmo não tenha sido observado para novilhas. O objetivo do trabalho foi desenvolver um modelo capaz de estimar a produção de leite do primeiro controle de vacas primíparas (P_LEITE) com base em variáveis rotineiramente coletadas nos controles leiteiros, de tal forma que pudesse ser utilizado na avaliação do período de transição de animais nessa categoria. Para tal, foram utilizadas informações do primeiro controle de 3267 primíparas da raça Holandesa, realizados entre 2014 e 2017. As informações foram alocadas em arquivos de treino ($n = 2345$) e validação ($n = 780$), posteriormente utilizados na estimativa dos parâmetros e avaliação dos modelos, respectivamente. Variáveis relacionadas com a P_LEITE foram identificadas através da *regularized regression* por meio da *elastic net*. Três tipos de modelos foram avaliados: regressão linear múltipla (RLM), *random forest* (RF) e rede neural artificial (RNA). Os modelos resultantes foram avaliados com base em seis estatísticas de ajuste e *10-fold cross validation*. Além disso, foram calculados os coeficientes de correlação de Pearson (r) e de concordância de Lin (ρ) entre os valores observados e preditos pelos modelos, assim como a comparação do valor mediano observado e predito de P_LEITE, adotando 5% como nível de significância. Os três modelos avaliados apresentaram capacidade preditiva para a P_LEITE, com erro absoluto médio abaixo de 4,00 kg no arquivo de validação e *10-fold cross validation*. Não foi observada diferença significativa entre os valores medianos preditos e observados. A RNA foi o modelo que apresentou a melhor qualidade de ajuste comparada com a RLM e RF. Trabalhos futuros poderão estudar as possíveis variações existentes entre rebanhos, como forma de melhorar a capacidade preditiva da RNA.

Palavras-chave: Aprendizado de máquina. Distúrbios metabólicos. Novilhas. Zootecnia de precisão.

ABSTRACT

Profound physiological and morphological changes occur in the body of a dairy cow during the transition period. Most of the metabolic disorders and infectious diseases are diagnosed during this time compromising the lactation overall success. Even though it has been proposed methods to evaluate the transition program in multiparous dairy cows, the same has not been observed for dairy heifers. The present paper aimed at developing a model capable of predicting first test day milk yield (P_LEITE) of primiparous dairy cows using Dairy Herd Improvement metrics (DHI). We used information from the first test day of 3267 Holstein primiparous dairy cows collected between 2014 and 2017. Data were split into training ($n = 2345$) and validation ($n = 780$) data set that were used to estimate model parameters and to evaluate the models, respectively. Variables associated to P_LEITE were identified using regularized regression via elastic net. Three types of models were evaluated: multivariate linear regression (MLR), random forest (RF) and artificial neural network (ANN). The resulting models were evaluated based on six fit-statistics and 10-fold cross validation. In addition, Pearson correlation coefficient (r) and Lin's concordance correlation coefficient (ρ) were calculated between predicted and observed values, as well as the comparison between observed and predicted median P_LEITE. Statistical significance was declared at an error level $\alpha < 0.05$. The three models were capable of predicting P_LEITE with a mean absolute error lower than 4.00 kg on validation data set and 10-fold cross validation. There was no significant difference between the observed and predicted median P_LEITE. ANN better-predicted P_LEITE compared to MLR and RF. Future work could focus on studying the variation between herds as a possibility to improve the ANN model.

Keywords: Dairy heifers. Machine learning. Metabolic disorders. Precision dairy farming

SUMÁRIO

CAPÍTULO 1 – REFERENCIAL TEÓRICO	13
1 INTRODUÇÃO	13
2 PERÍODO DE TRANSIÇÃO	14
3 DISTÚRBIOS METABÓLICOS E DOENÇAS INFECCIOSAS	17
4 EFEITO DO PERÍODO DE TRANSIÇÃO NA PRODUÇÃO ANIMAL	21
5 FORMAS DE AVALIAR O PERÍODO DE TRANSIÇÃO	22
6 CONSIDERAÇÕES FINAIS	25
CAPÍTULO 2 – PREVENDO A PRODUÇÃO DE LEITE DO PRIMEIRO CONTROLE LEITEIRO DE PRIMÍPARAS	27
1 INTRODUÇÃO	27
2 MATERIAL E MÉTODOS	28
2.1 Criação do arquivo de trabalho	29
2.2 Seleção de variáveis	30
2.3 Modelagens avaliadas	32
2.3.1 Regressão linear multivariada	32
2.3.2 Random forest	33
2.3.3 Rede neural artificial	33
2.4 Avaliação dos modelos	36
3 RESULTADOS	38
4 DISCUSSÃO	43
5 CONCLUSÃO	45
REFERÊNCIAS	47

CAPÍTULO 1 – REFERENCIAL TEÓRICO

1 INTRODUÇÃO

A atividade leiteira, assim como as demais atividades pecuárias, apresenta característica multifatorial, ou seja, inúmeras características e fatores se relacionam, ditando o resultado final. Dessa forma, produtores e técnicos envolvidos no manejo diário dos animais estão em constante processo de tomada de decisões. Decisões essas que quando tomadas pontualmente, visando melhorias específicas em determinado fator produtivo, podem impactar negativamente outros fatores, como por exemplo, a tendência antagônica observada entre intensificação da produção e mortalidade dos animais (PINEDO; DE VRIES; WEBB, 2010), assim como entre a produção de leite e o desempenho reprodutivo de vacas leiteiras (CHAGAS *et al.*, 2007). Além dessas, é possível que outras relações tanto favoráveis quanto desfavoráveis existam, mesmo que ainda não tenham sido identificadas.

Tendo como referência o ciclo produtivo de vacas leiteiras, a transição entre gestante e não lactante, para não gestante e lactante é denominado como período de transição (DRACKLEY, 1999). Alterações anatômicas, fisiológicas, metabólicas e hormonais, frequentemente observadas nesse período, são acompanhadas do balanço energético negativo (MANN *et al.*, 2015) e disfunção do sistema imunológico do animal (ROCHE *et al.*, 2013; STER; LOISELLE; LACASSE, 2012). Isso torna o mesmo susceptível ao desenvolvimento de doenças e distúrbios metabólicos (ØSTERGAARD; GRÖHN, 1999) que, por sua vez, irão comprometer o seu desempenho produtivo (RODRÍGUEZ; ARÍS; BACH, 2017; SANTSCHI *et al.*, 2016) e reprodutivo (CHAPINAL *et al.*, 2012; SEIFI; KIA, 2017). Dessa forma, é necessário o desenvolvimento de ferramentas capazes de avaliar efetivamente o período de transição de vacas leiteiras.

Uma vez que são rotineiramente coletadas, o desenvolvimento de tecnologias utilizando variáveis obtidas a partir de controles leiteiros oficiais (DHI) é encorajado, já que não é necessário o despendimento de recursos para a implementação de nova rotina de coleta de variáveis. Além disso, recentes avanços computacionais facilitaram a implementação de técnicas de modelagem mais complexas, como por exemplo *random forest* (BREIMAN, 2001a) e redes neurais artificiais (BENGIO, 2012). Ambas as técnicas, em especial a última, têm apresentado resultados animadores quando se trabalha com grande número de observações e variáveis, características comumente observadas em banco de dados de DHI (BROTZMAN *et al.*, 2015).

Ademais, informações de DHI têm se mostrado promissoras na determinação da incidência de distúrbios metabólicos, como cetose e a sua influência na composição do leite (SANTSCHI *et al.*, 2016), além de possibilitar a caracterização multivariada de diferentes rebanhos leiteiros (BROTZMAN *et al.*, 2015; DALLAGO *et al.*, 2017) e a avaliação do período de transição de vacas multíparas (NORDLUND, 2006).

O presente capítulo teve como objetivo revisar a literatura atualmente disponível sobre o período de transição de vacas leiteiras, abordando a sua definição, as alterações fisiológicas frequentemente observadas, a prevalência de distúrbios metabólicos e as formas de avaliação disponíveis.

2 PERÍODO DE TRANSIÇÃO

O período de transição se inicia três semanas antes do parto e se estende por até três semanas após o parto (DRACKLEY, 1999; GRUMMER, 1995), sendo marcado por profundas alterações anatômicas, fisiológicas, metabólicas e hormonais no corpo da vaca leiteira (PONCHEKI, 2015) e impondo elevado desafio metabólico, especialmente para animais de alta produtividade (ROCHE *et al.*, 2013). Durante esse período, é frequentemente observada a redução na capacidade do animal em ingerir alimentos. Segundo Grummer, Mashek e Hayirli (2004) ocorre decréscimo no consumo de matéria seca de aproximadamente 30 e 24%, respectivamente para multíparas e primíparas, ao longo do período de transição no pré-parto.

Concomitantemente à redução no consumo de alimentos, é observado aumento abrupto na demanda nutricional devido ao início da produção de leite (BELL, 1995). Segundo Bell (1995), as exigências por aminoácidos e glicose de um animal produzindo cerca de 30 kg de leite por dia, no quarto dia pós-parto é, respectivamente, 2 e 2,7 vezes maior quando comparado com a exigência no final da gestação. Já Reynolds *et al.* (2003) estimaram aumento de 50% na exigência de glicose em animais no 11º dia pós-parto, quando comparada à exigência de vacas no nono dia pré-parto. Além disso, o teor de minerais presente no colostro, principalmente de cálcio e fósforo, é maior do que aquele observado no leite (MCGRATH *et al.*, 2016), indicando maior exigência mineral pelos animais durante o período de transição.

Em consequência ao aumento da exigência nutricional que não é acompanhada proporcionalmente por aumento no consumo de alimentos (BAUMAN; CURRIE, 1980), vacas leiteiras encontram-se em estado de balanço energético negativo (BEN) durante o período de transição. Mann *et al.* (2015) avaliaram três diferentes estratégias nutricionais no período seco, diferindo no conteúdo de energia, amido e fibra e, independente do tratamento, foi observado

BEN entre a primeira e a sexta semana pós-parto, em animais com produção de leite acima de 40kg por dia. Observa-se também a redução na concentração sérica de insulina (FIORE *et al.*, 2018), já que esse é o principal hormônio estimulador da lipogênese, ao passo que aumenta a concentração de hormônios que estimulam a lipólise, como o hormônio do crescimento e o glucagon (SANTOS, 2011). Dessa forma, ocorre a mobilização de reservas corporais numa tentativa de atender à exigência do animal. Conseqüentemente, ocorre redução significativa no escore de condição corporal nas primeiras semanas de lactação e esse efeito é ainda mais acentuado em animais que pariram com grande quantidade de reserva (GHEISE *et al.*, 2017; HEUER; SCHUKKEN; DOBBELAAR, 1999).

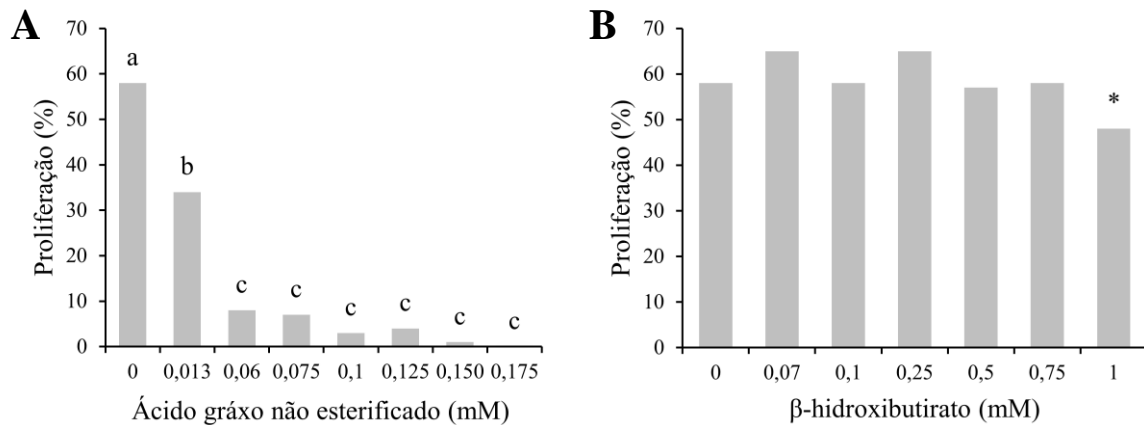
Em resposta à mobilização de reservas, é observado aumento na concentração sanguínea de ácidos graxos não esterificados (AGNE) e β -hidroxibutirato (BHB) (BARLETTA *et al.*, 2017; GHEISE *et al.*, 2017; REYNOLDS *et al.*, 2003). Em condições de BEN extremo, a elevada mobilização e conseqüente captação hepática de ácidos graxos, é maior que a capacidade de metabolização do órgão, podendo acarretar o surgimento de esteatose hepática, já que o fígado de ruminantes apresenta reduzida capacidade de exportar triacilgliceróis por meio de lipoproteínas de densidade muito baixa e então, acabam se acumulando no parênquima hepático (SANTOS, 2011).

A completa metabolização hepática de ácidos graxos requer o seu transporte para dentro do espaço mitocondrial, por meio da ativação do complexo enzimático carnitina-palmitoil transferase (PALMQUIST; MATTOS, 2011), o que foi observado em vacas em transição (DOUGLAS *et al.*, 2006). Grande quantidade de acetato resulta da completa oxidação de ácidos graxos, metabólito esse que pode ser desviado para a síntese de acetoacetato, sendo esse o principal corpo cetônico resultante do metabolismo hepático (PALMQUIST; MATTOS, 2011). Entretanto, como esse é rapidamente descarboxilado a acetona, e essa é em grande parte eliminada durante a expiração e confere o odor característico de acetona ao hálito, o BHB é o corpo cetônico mais estável na circulação sanguínea dos ruminantes (SANTOS, 2011).

Excessiva quantidade de AGNE no plasma sanguíneo de bovinos tem sido relacionada com a disfunção do sistema imune de animais em transição. Em trabalho realizado por Ster, Loisel e Lacasse (2012), foi observada correlação negativa entre a concentração de células mononucleares do sangue periférico (CMSP) e AGNE ($r = -0,86$; $P < 0,001$), assim como entre CMSP e BHB ($r = -0,54$; $P < 0,01$). Os mesmos autores conduziram posteriormente o cultivo *in vitro* de CMSP com diferentes doses de inclusão de AGNE e BHB no meio de cultura, sendo constatado decréscimo drástico na proliferação dessas células de defesa com a inclusão de AGNE (Figura 1 – A), não sendo o mesmo observado com a inclusão de BHB

(Figura 1 – B). Efeito similar também foi encontrado por Lacetera *et al.* (2004), que observaram redução tanto na proliferação de CMSP e respectiva secreção de interferon gama, quanto na secreção de imunoglobulinas M ao adicionar níveis crescente de AGNE no meio de cultura.

Figura 1 – Efeito de níveis crescentes de ácidos graxos não esterificados (A) e β -hidroxibutirato (B) na proliferação *in vitro* de células mononucleares do sangue periférico. Letras diferentes na figura A indicam diferença estatisticamente significativa ($P < 0,001$), enquanto que na figura B * indica diferença significativa em relação ao controle (0 mM; $P < 0,05$).



Fonte: Ster, Loisel e Lacasse (2012). Adaptado.

A funcionalidade de leucócitos periféricos também é prejudicada com altas concentrações de AGNE. Em trabalho conduzido por Hammon *et al.* (2006), foi encontrado que a atividade de leucócitos periféricos foi reduzida em animais que apresentaram elevada concentração sérica de AGNE. Além disso, resultados *in vitro* indicam que a capacidade oxidativa e de realizar fagocitose de leucócitos periféricos é reduzida em altas concentrações de AGNE (SCALIA *et al.*, 2006), o que explicaria em parte, a reduzida capacidade de neutrófilos de animais na primeira semana pós-parto em combater células bacterianas (KEHRLI; NONNECKE; ROTH, 1989).

Esses resultados indicam efeito negativo de AGNE em células do sistema imunológico, explicando assim a disfunção imune observada em vacas durante o período de transição.

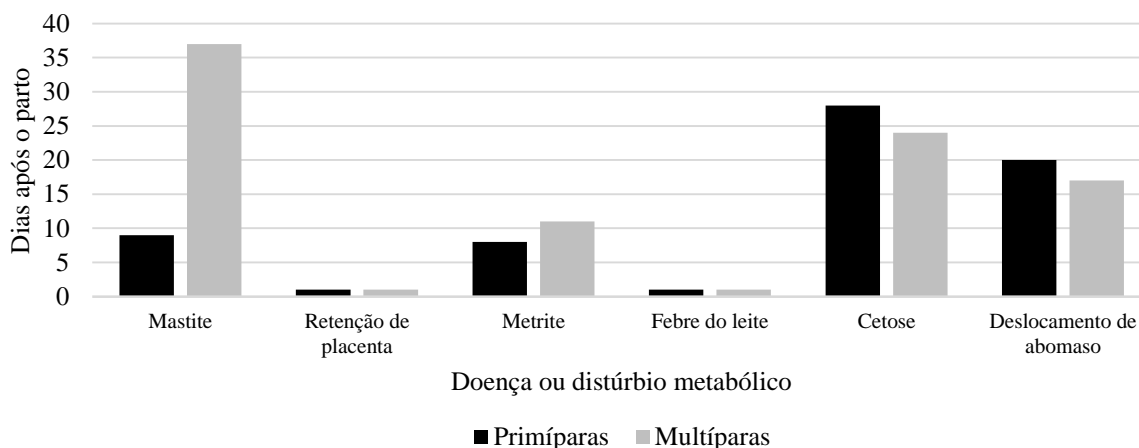
Animais com maior capacidade de produzir leite estariam mais sujeitos aos efeitos negativos da concentração sanguínea elevada de AGNE e BHB. Quanto maior for a produção de leite ao início da lactação, maior será a mobilização de tecido adiposo para atender a demanda energética e garantir a produção leiteira (LACASSE *et al.*, 2018). A concentração sérica de AGNE e BHB será proporcional, impactando negativamente o sistema imunológico desses animais.

A produção de leite tanto de vacas multíparas quanto primíparas aumenta de acordo com a frequência de ordenha (BARNES; PEARSON; LUKES-WILSON, 1990). Portanto, uma possível estratégia proposta a fim de amenizar o estresse metabólico do período de transição, seria reduzir o número de ordenhas diárias nas semanas logo após o parto, o que consequentemente reduziria a produção de leite, o BEN, a concentração sérica de AGNE e BHB e, melhoraria a capacidade imune dos animais durante a transição. Em trabalho realizado por Loiselle *et al.* (2009) foram avaliadas as concentrações séricas de AGNE e BHB, assim como a produção de leite de vacas multíparas ordenhadas uma vez por dia, na primeira semana pós-parto e duas vezes por dia nas demais semanas da lactação, comparada a vacas que foram ordenhas duas vezes por dia durante todo o período experimental. Nos animais ordenhados duas vezes ao dia foi observada maior produção de leite na primeira semana, mas a partir da segunda semana, essa diferença não foi mais significativa. Por outro lado, as concentrações de AGNE e BHB foram menores até 24 e 14 dias pós-parto, respectivamente, para animais ordenhados uma vez por dia, sugerindo que a redução no estímulo para a produção de leite na primeira semana pós-parto, melhorou o perfil metabólico dos animais. Essa estratégia se mostraria ainda mais interessante para vacas obesas no pré-parto, já que essas são mais propensas a desenvolver distúrbios metabólicos como esteatose hepática e cetose, devido à maior mobilização de tecido adiposo (GHEISE *et al.*, 2017).

3 DISTÚRBIOS METABÓLICOS E DOENÇAS INFECCIOSAS

Doenças infecciosas e distúrbios metabólicos são, em sua maioria, diagnosticados durante o período de transição pós-parto. Segundo LeBlanc *et al.* (2006), cerca de 75% das doenças que acometem o gado leiteiro são diagnosticadas durante o primeiro mês de lactação (Figura 2). Além disso, entre 30 e 50% dos animais são afetados por algum tipo de distúrbio durante a transição (LEBLANC, 2010). Dessa forma, o período de transição é frequentemente desastroso para a vaca leiteira (GOFF; HORST, 1997), podendo ditar a rentabilidade de toda a lactação (DRACKLEY, 1999).

Figura 2 – Dia mediano do diagnóstico de algumas doenças e distúrbios metabólicos que acometem o gado leiteiro em relação ao dia do parto.



Fonte: Østergaard e Gröhn (1999). Adaptado.

A prevalência de diferentes distúrbios metabólicos durante o período de transição é variada. Suthar *et al.* (2013) avaliaram a ocorrência de diversos distúrbios nos primeiros 30 dias pós-parto, de 5884 animais criados em 528 rebanhos pertencentes a 10 países europeus. Eles encontraram maior prevalência de cetose subclínica (21,8%) seguida por retenção de placenta (10,4%) e metrite (9,6). A prevalência de cetose subclínica variou entre 26,4 e 55,7% em vacas confinadas em galpões do tipo *free-stall* e avaliadas até o 16º dia pós-parto (MCART *et al.*, 2011). Santschi *et al.* (2016) avaliaram a prevalência de cetose em rebanhos canadenses e encontraram prevalência média de 22,9%, sendo o maior número de casos em primíparas, observados na primeira semana pós-parto (30%).

Daros *et al.* (2017) avaliaram a incidência de diversos distúrbios metabólicos e doenças durante o período de transição de vacas leiteiras mantidas a pasto, em 53 propriedades localizadas no estado de Santa Catarina. Entre as condições avaliadas, a cetose subclínica apresentou o maior número de casos seguida por retenção de placenta e metrite. Por outro lado, Ribeiro *et al.* (2013) encontraram hipocalcemia subclínica como sendo a condição mais prevalente em animais também mantidos em pastejo, seguida por cetose subclínica, diferindo do observado por Sepúlveda-Varas *et al.* (2015), ao avaliarem animais mantidos a pasto em seis rebanhos chilenos, onde a doença mais prevalente foi metrite leve, seguida por hipocalcemia e cetose, ambas subclínicas.

Cetose subclínica tem sido apresentada por alguns autores como sendo um dos principais fatores de risco para o desenvolvimento de outros problemas, devido à sua relação com o BEN (GRÖHN *et al.*, 1989; SUTHAR *et al.*, 2013). Animais que desenvolvem cetose subclínica na primeira semana pós-parto são mais propensos a desenvolverem problemas de

saúde (MCART; NYDAM; OETZEL, 2012) como endometrite, ovário cístico (SHIN *et al.*, 2015), deslocamento de abomaso (GEISHAUSER *et al.*, 1997) e laminite (HEUER; SCHUKKEN; DOBBELAAR, 1999). Além de estar negativamente associada com a capacidade dos animais de emprenharem na primeira inseminação (RUTHERFORD; OIKONOMOU; SMITH, 2016).

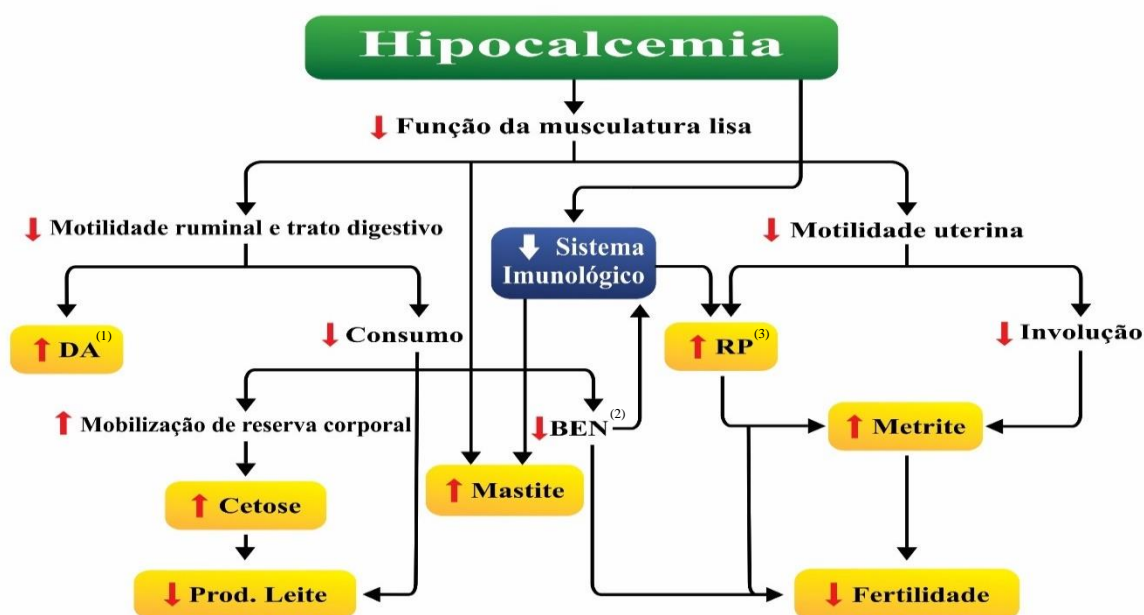
Foram propostas diferentes maneiras e limiares na identificação do estado subclínico e clínico de cetose. A mensuração dos teores de BHB sanguíneos tem sido a forma mais aceita (OETZEL, 2004), já que esse é o corpo cetônico predominante (SANTOS, 2011). Frequentemente, o valor limiar adotado de BHB na identificação de cetose é de 12,48 mg dL⁻¹ ou 1,2 mMol L⁻¹ (SANTSCHI *et al.*, 2016), sendo o mesmo recomendado para animais mantidos a pasto (COMPTON; YOUNG; MCDUGALL, 2015). Uma alternativa menos invasiva e mais prática, seria por meio da determinação de BHB no leite, adotando limiares de < 1,56 mg/dL (< 0,15 mMol L⁻¹) e ≥ 2,08 mg dL⁻¹ (≥ 0,20 mMol L⁻¹) para a identificação, respectivamente, de animais não acometidos e acometidos, tendo os valores intermediários como animais suspeitos (DENIS-ROBICHAUD *et al.*, 2014).

Além disso, o monitoramento do escore de condição corporal (ECC) durante o período de transição seria uma alternativa prática na identificação de animais possivelmente acometidos com cetose, já que animais que parem com ECC maior ou igual a 3,5 são mais propensos a desenvolverem cetose (GILLUND *et al.*, 2001). Uma das principais vantagens dessa alternativa está na ausência de custos associados, assim como, a sua facilidade de emprego.

Mais recentemente, observou-se que a hipocalcemia subclínica foi o distúrbio metabólico de maior incidência durante o período de transição. Reinhardt *et al.* (2011) avaliaram a concentração sérica de cálcio, em amostras coletadas 48 horas pós-parto, de 1462 animais de 480 propriedades produtoras de leite de 21 estados dos Estados Unidos da América. Eles identificaram prevalência de hipocalcemia clínica de 5% ao passo que a hipocalcemia subclínica variou entre 41 e 54% nos animais entre a 2^o e 5^o lactação, respectivamente, indo ao encontro com a prevalência de 43,3% observada por Ribeiro *et al.* (2013).

Argumenta-se então que a hipocalcemia seja o principal distúrbio metabólico que acomete o gado leiteiro durante o período de transição (SEIFI; KIA, 2017). Além disso, avanços nas pesquisas avaliando a hipocalcemia em vacas no periparto, têm sugerido que esse distúrbio seja responsável por desencadear reações em cascata, que irão propiciar o desenvolvimento das demais doenças e distúrbios metabólicos indesejados durante o período de transição (GOFF, 2008; WEAVER *et al.*, 2016) (Figura 3).

Figura 3 – Representação esquemática dos efeitos da hipocalcemia em vacas leiteiras.



Fonte: Howard e Smith (1999) citado por Seifi e Kia (2017). Adaptado.

(¹) Deslocamento de abomaso. (²) Balanço energético negativo. (³) Retenção de placenta.

O cálcio é essencial para inúmeros processos fisiológicos tais como, contração muscular, transmissão nervosa e segundo mensageiro na modulação de respostas hormonais, além de ser o principal constituinte da matriz óssea (SEIFI; KIA, 2017). A redução nos níveis séricos desse mineral está negativamente associada com a produção de leite, assim como, o desempenho reprodutivo (ROCHE *et al.*, 2013). Multíparas são mais propensas a apresentarem hipocalcemia logo após o parto (REINHARDT *et al.*, 2011), muito provavelmente em resposta ao maior volume de colostro produzido quando comparado a primíparas (VANHOLDER *et al.*, 2015), e pela quantidade elevada desse mineral em sua composição (MCGRATH *et al.*, 2016).

Dessa forma, mecanismos homeostáticos responsáveis por regular a concentração sérica de cálcio são ativados logo após o parto, ocorrendo estímulo hormonal visando aumentar a reabsorção de cálcio nos rins, aumentar a absorção de cálcio dietético e promover a reabsorção óssea (GOFF, 2008).

O cálcio também é essencial na atividade do sistema imunológico, modulando a capacidade fagocitária de neutrófilos. Durante o pós-parto, é reduzida a concentração intracelular de cálcio em células mononucleares do sangue periférico, essa redução é ainda mais proeminentes em animais que desenvolvem hipocalcemia clínica (KIMURA; REINHARDT; GOFF, 2006). Também, é observada redução no número de células e a sua funcionalidade mesmo em animais com hipocalcemia subclínica (MARTINEZ *et al.*, 2012). Juntos, esses resultados explicariam em parte, a maior incidência de doenças infecciosas como metrite e mastite durante

o período de transição (Figura 3). Além disso, animais que desenvolveram hipocalcemia subclínica entre 24 e 48 horas pós-parto foram 3,71; 5,47; 4,25 e 3,43 vezes mais propensos a desenvolverem deslocamento de abomaso, cetose clínica, metrite e retenção de placenta, respectivamente, no trabalho conduzido por Rodríguez; Arís e Bach (2017), estando também associado com o descarte do animal até 60 dias em lactação, conforme Seifi *et al.* (2011).

A identificação de animais acometidos por hipocalcemia se dá pela determinação dos níveis séricos de cálcio. Níveis normais variam entre 8 e 10 mg/dL (2 e 2,5 mMol/L) (WEAVER *et al.*, 2016). Valores entre 5,6 e 8 mg/dL (1,4 a 2 mMol/L) indicam hipocalcemia subclínica, ao passo que, valores iguais ou menores que 5,6 mg/dL (1,4 mMol/L) indicam hipocalcemia clínica (GOFF, 2008; REINHARDT *et al.*, 2011).

4 EFEITO DO PERÍODO DE TRANSIÇÃO NA PRODUÇÃO ANIMAL

Alterações na produção de leite é uma das principais consequências do período de transição. Elevadas concentrações de AGNE e BHB no sangue durante as duas primeiras semanas de lactação, juntamente com baixa concentração sérica de cálcio durante as três primeiras semanas de gestação, estão relacionados com a redução na produção de leite no primeiro controle leiteiro (CHAPINAL *et al.*, 2012).

Tanto a quantidade quanto a composição química do leite são alteradas em animais com cetose. A produção de leite no primeiro controle de animais diagnosticados com cetose tende a ser menor do que aquela observada em animais saudáveis (BENEDET *et al.*, 2017; SANTSCHI *et al.*, 2016), efeito negativo que se estende ao longo de toda a lactação (RAJALA-SCHULTZ; GRÖHN; MCCULLOCH, 1999). Por outro lado, é comum o aumento do teor de sólidos totais no leite de animais acometidos por esse distúrbio, sendo a alteração ocasionada principalmente pelo aumento no teor de gordura (BENEDET *et al.*, 2017; SANTSCHI *et al.*, 2016).

Além disso, vacas com elevado BEN e que desenvolvem cetose apresentam, consequentemente, menor eficiência reprodutiva. Em trabalho conduzido por Garverick *et al.* (2013), foi encontrado que vacas que ficaram prenhas na primeira inseminação, realizada 75 ± 7 dias pós-parto, tiveram concentração sérica de AGNE aos 3, 7 e 14 dias pós-parto, menores que vacas que não ficaram prenhas ($P < 0,05$). A concentração sérica de AGNE, juntamente com BHB, são comumente utilizadas como indicativo do estado energético de vacas leiteiras no início da lactação (BEEVER, 2006). Dessa forma, os resultados de Garverick *et al.* (2013) sugerem que animais com melhor estado energético ficariam prenhas mais cedo, reduzindo o

intervalo entre partos, fator esse que tem sido apresentado como um dos mais importantes para maximizar o retorno financeiro da atividade leiteira (INCHAISRI *et al.*, 2010).

Além do exposto, seria interessante mencionar possíveis efeitos de variáveis ambientais sobre o animal em transição. Em condições de estresse térmico por calor, vacas leiteiras tendem a reduzir significativamente o seu consumo de matéria seca ao passo que aumentam proporcionalmente o consumo de água (WEST, 2003). Conseqüentemente, ocorre redução significativa na produção de leite, além de ser frequentemente observada redução no desempenho reprodutivo dos animais em condição de estresse por calor (RENSIS; SCARAMUZZI, 2003). Tais efeitos possuem magnitude ainda maior em animais em transição, somando-se às condições negativas que já são frequentemente observadas.

5 FORMAS DE AVALIAR O PERÍODO DE TRANSIÇÃO

Frente ao apresentado nos tópicos anteriores, a redução no consumo de alimentos e aumento nos níveis séricos de AGNE e BHB, assim como a relativa maior incidência de doenças e distúrbios metabólicos e hipocalcemia são os principais eventos observados durante o período de transição de vacas leiteiras. Uma possível abordagem na tentativa de avaliar a transição deveria levar em consideração tais eventos. Embora a mensuração desses eventos seja relativamente fácil em condições experimentais, a sua mensuração em condições frequentemente encontradas em propriedades produtoras de leite, apresenta limitações consideráveis.

A avaliação da incidência de doenças e distúrbios metabólicos seria uma abordagem direta na avaliação do período de transição. Entretanto, embora a incidência seja maior no início da lactação (ØSTERGAARD; GRÖHN, 1999), a frequência com que eles ocorrem é relativamente baixa (DAROS *et al.*, 2017; JORDAN; FOURDRAINE, 1993; VAN DORP *et al.*, 1999). Além disso, existe grande variação na capacidade de diagnóstico pelo produtor rural, definição e terminologia dos distúrbios e doenças, além de baixa disciplina no registro do acontecimento de distúrbios metabólicos e doenças por parte dos produtores rurais (NORDLUND; COOK, 2004). Conseqüentemente, é dificultado o uso da incidência desses problemas como forma de avaliar efetivamente o período de transição.

A redução na ingestão de alimentos tem sido o evento observado com maior frequência na transição pós-parto (BAUMAN; CURRIE, 1980; GRUMMER; MASHEK; HAYIRLI, 2004; MANN *et al.*, 2015). Dessa forma, monitorar a ingestão de alimentos e sua respectiva redução poderia auxiliar na avaliação do período de transição. De fato, o monitoramento desse

evento se mostrou eficaz em identificar com pelo menos um dia de antecedência, animais que seriam diagnosticados com cetose e laminite (GONZÁLEZ *et al.*, 2008). Por outro lado, o comportamento ingestivo de múltiparas e primíparas não é o mesmo (NEAVE *et al.*, 2017), indicando a necessidade de considerar o fator paridade. Entretanto, a mensuração do consumo de alimentos por cada animal em condições comerciais é restrita, dificultando a sua efetiva utilização.

De forma similar, a mensuração sérica de AGNE e BHB também apresenta limitações práticas. Alterações fisiológicas no corpo da vaca em transição podem ser eficientemente representadas por mudanças nos teores séricos de AGNE e BHB (SHRESTHA, 2015). Em escala comercial, existem testes que podem ser utilizados na avaliação da concentração de BHB no sangue apresentado boa acurácia na determinação (PINEDA; CARDOSO, 2015). Entretanto, a utilização desse tipo de teste requer tempo hábil para a coleta de amostras de sangue em cada animal, podendo ser difícil em grandes rebanhos, além de aumentar o custo de produção.

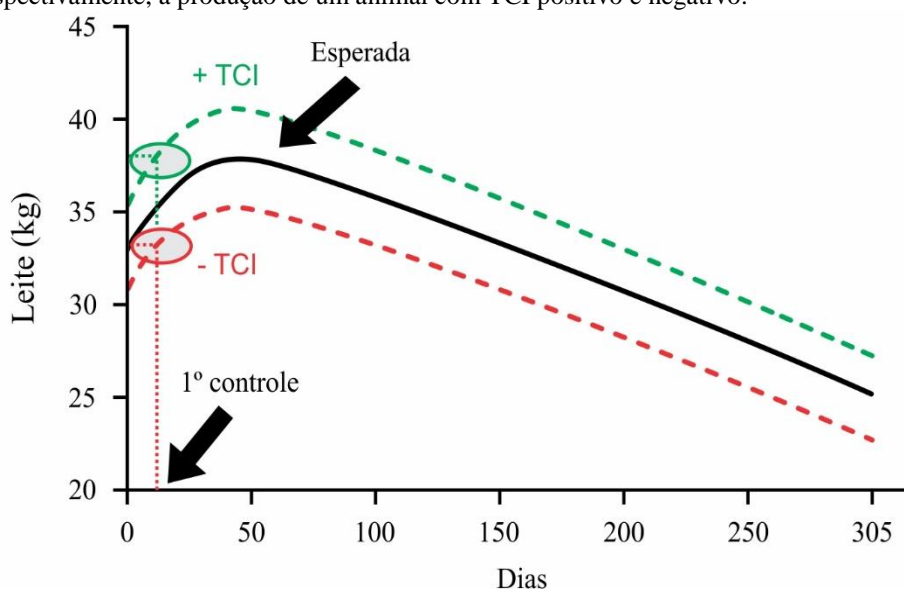
Como alternativa ao sangue, constituintes do leite poderiam ser utilizados na avaliação do período de transição. Possível alternativa seria a avaliação de BHB no leite, uma vez que amostras são rotineiramente coletadas de todos os animais em lactação em propriedades que participam do controle leiteiro realizado por associações. Embora a concentração de BHB no leite seja menor que aquela observada no plasma sanguíneo (SANTOS, 2011), existe alta correlação entre BHB sanguíneo e do leite [$r = 0,89$; (DENIS-ROBICHAUD *et al.*, 2014)].

Além disso, a relação entre os teores de gordura e proteína do leite (FPR) poderia auxiliar na avaliação do período de transição. O teor de gordura do leite de vacas em BEN tende a ser elevado (BENEDET *et al.*, 2017; SANTSCHI *et al.*, 2016), dada a mobilização e consecutiva disponibilidade de ácidos graxos na corrente sanguínea que são captados pela glândula mamária e, posteriormente, utilizados na síntese da gordura do leite (PALMQUIST; MATTOS, 2011). Alta correlação tem sido apresentada entre BHB no leite e FPR [$r = 0,89$; $P < 0,001$ (SANTSCHI *et al.*, 2016)], muito embora a correlação entre BHB sanguíneo e FPR seja baixa [$r = 0,17$; $P = 0,03$ (DENIS-ROBICHAUD *et al.*, 2014)].

O descarte prematuro de animais está relacionado com a ocorrência de doenças no início da lactação (PINEDO; DE VRIES; WEBB, 2010), além de estar relacionado com a concentração sérica de cálcio, AGNE e BHB entre uma semana pré-parto até duas semanas pós-parto (ROBERTS *et al.*, 2012). Dessa forma, foi proposto utilizar a taxa de descarte de animais até 60 dias em lactação como possível forma de avaliar o manejo de transição adotado na propriedade (NORDLUND; COOK, 2004).

Devido à relação negativa observada entre a ocorrência de problemas e a produção de leite durante o período de transição (BENEDET *et al.*, 2017; EDWARDS; TOZER, 2004; HUZZEY *et al.*, 2007; PROUDFOOT; HUZZEY; VON KEYSERLINGK, 2009; RAJALA-SCHULTZ; GRÖHN; MCCULLOCH, 1999; SANTSCHI *et al.*, 2016), essa variável foi proposta como sendo o principal reflexo do manejo de transição empregado, podendo ser então utilizado na sua avaliação. Partindo dessa premissa, foi desenvolvido o *Transition Cow Index*TM [TCI (NORDLUND, 2006)] com a finalidade de avaliar o resultado do período de transição de vacas leiteiras (Figura 4).

Figura 4 – Representação da curva de lactação, indicando a relação entre valores positivos e negativos do *Transition Cow Index*TM (NORDLUND, 2006) com a produção de leite em 305 dias. A linha contínua indica a produção de leite esperada, ao passo que as linhas tracejadas em verde e vermelho representam, respectivamente, a produção de um animal com TCI positivo e negativo.



O modelo patenteado do TCI estima a produção de leite no primeiro controle leiteiro de uma nova lactação, realizado entre 5 e 40 dias pós-parto, utilizando variáveis obtidas no próprio controle assim como informações do último controle da lactação anterior (NORDLUND *et al.*, 2011). O resultado da predição pode então ser utilizado pela agência ou associação de criadores, responsáveis por conduzir o controle leiteiro, para estimar a curva de produção em 305 dias de lactação. De forma análoga, a produção de leite observada no primeiro controle é também utilizada para estimar a curva de produção. O TCI é então obtido pela diferença entre a produção de leite observada e predita. Ele pode tanto ser apresentado como a diferença no primeiro controle leiteiro, assim como a diferença em 305 dias de lactação [Figura 4 (SCHULTZ *et al.*, 2016)]. Independentemente da forma como o resultado é expresso, busca-se valores positivos de TCI, pois esses indicam que o período de transição foi bem sucedido, sendo o contrário verdadeiro (NORDLUND, 2006).

Lukas *et al.* (2015) avaliaram a taxa de incremento na produção de leite nos primeiros sete dias de lactação, a produção de leite média dos sete dias e a diferença entre a produção média observada e estimada, como possíveis indicadores para avaliar o período de transição. Os melhores resultados foram encontrados a partir da terceira abordagem que foi similar à forma como o TCI é obtido. Entretanto, o TCI apresenta maior facilidade de obtenção quando comparado à metodologia proposta por Lukas *et al.* (2015), uma vez que não é necessário coletar informações ao longo dos sete dias iniciais da lactação e utiliza variáveis rotineiramente coletadas durante controles leiteiros oficiais.

A ocorrência de distúrbios metabólicos como metrite, cetose e deslocamento de abomaso estão associados com valores negativos de TCI (NORDLUND, 2006). Além disso, os resultados do manejo de transição adotado em diferentes rebanhos podem ser comparados entre si com a utilização do TCI. Facilitando a comparação de diferentes manejos de transição adotados em diferentes propriedades, levando em consideração o efeito do rebanho.

6 CONSIDERAÇÕES FINAIS

O período de transição é particularmente complexo dada a relação multivariada simultânea de diferentes componentes da atividade leiteira, sendo portanto, um período de extrema importância, visando garantir o sucesso da lactação. Em multíparas, existem formas estabelecidas de se avaliar o manejo de transição adotado, sendo necessário o desenvolvimento de metodologia similar para primíparas.

CAPÍTULO 2 – PREVENDO A PRODUÇÃO DE LEITE AO PRIMEIRO CONTROLE LEITEIRO DE PRIMÍPARAS

1 INTRODUÇÃO

O período de transição do gado leiteiro se inicia três semanas antes do parto e se prolonga por até três semanas após o parto, sendo marcado por profundas alterações fisiológicas e endócrinas no corpo da vaca leiteira (DRACKLEY, 1999; GRUMMER, 1995). A redução no consumo de matéria seca tanto em primíparas quanto em multíparas é frequentemente observada durante esse período (GRUMMER; MASHEK; HAYIRLI, 2004), o que ocorre concomitantemente com o aumento das exigências nutricionais devido à crescente produção de leite na fase inicial da lactação (NRC, 2001).

O balanço energético negativo (BEN) e a disfunção do sistema imunológico são frequentemente observados durante o período de transição (DRACKLEY, 1999; DRACKLEY; CARDOSO, 2014; LACASSE *et al.*, 2018), tornando os animais propensos a desenvolverem distúrbios metabólicos, reduzirem a produção de leite, e apresentarem baixo desempenho reprodutivo futuro (CHAPINAL *et al.*, 2012; GANTNER; BOBIĆ; POTOČNIK, 2016; MCART; NYDAM; OETZEL, 2012; ZEBELI *et al.*, 2015). Dessa forma, o manejo adequado de animais em transição deve visar reduzir o BEN (LACASSE *et al.*, 2018), evitando que este afete de forma negativa e definitiva, o resultado final de toda a lactação (DRACKLEY, 1999).

Para primíparas, a transição entre o estado gestante não-lactante para não gestante e lactante, assume posição de destaque, uma vez que a produção de leite é homeorreticamente favorecida (DRACKLEY; CARDOSO, 2014), competindo por nutrientes outrora destinados ao crescimento corporal (NRC, 2001). Avaliando especificamente o período de transição em animais de primeira cria, Turk *et al.* (2013) reportaram a existência de um complexo sistema de adaptação ao BEN, com base na mobilização de tecido adiposo que se estabiliza com o avançar da lactação. Além disso, a ocorrência e severidade de acidose ruminal durante o período de transição de primíparas é aumentada logo após o parto, independentemente do tipo da dieta, controle ou alto teor de alimento concentrado (PENNER; BEAUCHEMIN; MUTSVANGWA, 2007). Esses resultados indicam que, durante o período de transição, tanto primíparas quanto multíparas enfrentam desafios metabólicos e alterações fisiológicas similares (DRACKLEY, 1999).

Torna-se evidente a importância do desenvolvimento de ferramentas capazes de avaliar efetivamente o manejo de animais em transição, que auxiliem técnicos e produtores

rurais na tomada de decisões gerenciais. Além disso, a aparente similaridade entre o metabolismo de transição de primíparas e múltíparas indica que uma ferramenta capaz de avaliar o período de transição de múltíparas seria, em teoria, eficaz na avaliação de primíparas.

O *Transition Cow Index*TM (TCI; NORDLUND, 2006) foi desenvolvido com a finalidade de avaliar o sucesso ou fracasso do período de transição de vacas leiteiras. Ele é obtido pela diferença entre a produção de leite observada e predita para o primeiro controle leiteiro, sendo o resultado extrapolado para 305 dias de lactação (SCHULTZ *et al.*, 2016). Entretanto, o TCI utiliza parâmetros obtidos tanto na lactação atual quanto passada para avaliar a transição atual (NORDLUND, 2006), não podendo, portanto, ser utilizado para primíparas.

Dessa forma, o presente trabalho teve como objetivo desenvolver um modelo utilizando informações disponíveis no primeiro controle leiteiro, que seja capaz de prever a produção de leite de vacas primíparas holandesas, podendo então ser utilizado na avaliação do período desses animais.

2 MATERIAL E MÉTODOS

Os dados utilizados no presente trabalho foram fornecidos pela empresa canadense Valacta – *Dairy Production Centre of Expertise in Québec and Atlantic Regions*, a partir de banco de dados pré-existente, não sendo, portanto necessária a aprovação do Comitê de Ética no Uso de Animais da Universidade Federal dos Vales dos Jequitinhonha e Mucuri.

Os dados foram validados removendo-se entradas inconsistentes e *outliers* (observações anômalas), e estimando-se parcelas perdidas. *Elastic net regularized regression* (regressão regularizada por meio da rede elástica) foi empregada na seleção de variáveis relacionadas com a produção de leite no primeiro controle (P_LEITE).

Posteriormente, modelos de regressão linear multivariada (RLM), *random forest* (RF) e rede neural artificial (RNA) foram avaliados quanto à sua capacidade preditiva em relação à variável resposta P_LEITE.

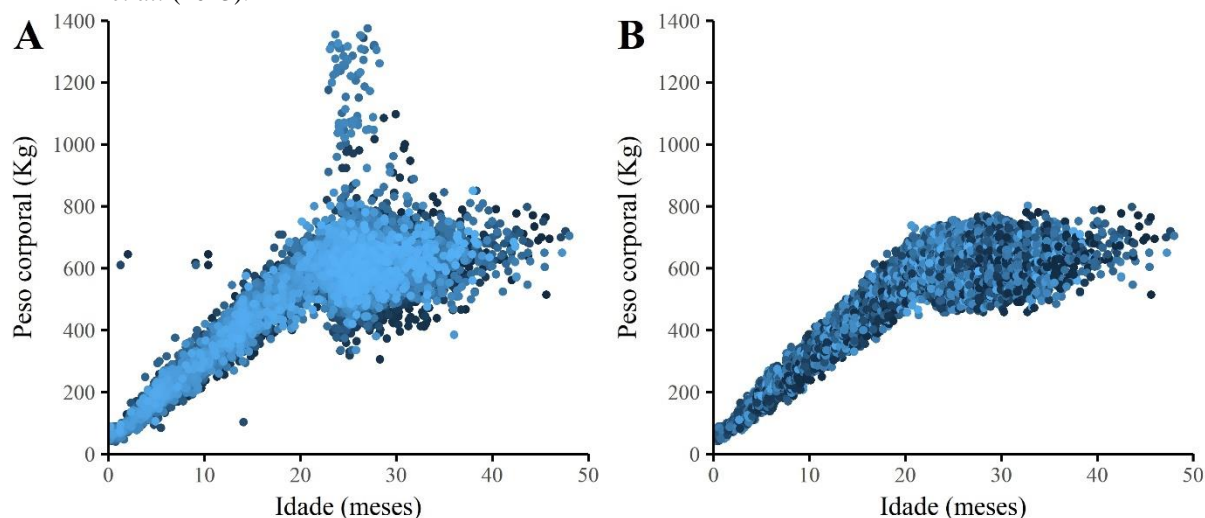
A preparação dos dados e a modelagem foram realizadas no programa estatístico R (R CORE TEAM, 2018), por meio de pacotes específicos, utilizando um computador com sistema operacional Windows 10 Edição Home (versão 64-bit), 6,00 GB de memória RAM e processador Intel Core i5 com 1,60 GHz.

2.1 Criação do arquivo de trabalho

As variáveis estavam contidas em três arquivos distintos denominados PESO, TESTE e LACTAÇÃO contendo, respectivamente, informações sobre o peso corporal, controle leiteiro e lactação completa de vacas primíparas da raça Holandesa, de 100 rebanhos localizados na província de Québec – Canadá. Com exceção do arquivo PESO, cada arquivo continha uma única observação por primípara, coletada no período de 2014 a 2017 em 8923 animais.

O arquivo PESO continha 11042 observações obtidas em 6478 animais entre 2009 e 2017. A identificação e eliminação de *outliers* nesse arquivo foram feitas seguindo metodologia proposta por Leys *et al.* (2013). Por essa metodologia, o intervalo de observações válidas foi definido como sendo 2,5 vezes o desvio absoluto da mediana dos pesos corporais, de acordo com a idade dos animais em meses (Figura 5).

Figura 5 – (A) Distribuição das observações de peso corporal em relação à idade dos animais apresentando *outliers*; (B) Distribuição das observações de peso corporal em relação à idade dos animais após a remoção de *outliers* identificados com base no desvio absoluto da mediana segundo metodologia proposta por Leys *et al.* (2013).



O modelo para animais da raça Holandesa desenvolvido por Cue *et al.* (2012), ao avaliar o crescimento de novilhas canadenses, foi utilizado para estimar o peso corporal de cada animal aos 3, 9, 12 e 15 meses de idade, assim como o ganho médio diário entre esses intervalos (VACEK *et al.*, 2015). O peso ao parto também foi estimado para cada animal utilizando o mesmo modelo.

O arquivo TESTE continha, inicialmente, 8923 observações coletadas entre janeiro de 2014 e outubro de 2017. Com o objetivo de manter apenas observações do primeiro controle leiteiro, foram descartadas observações nas quais a variável dias em lactação no controle era

superior a 40 dias ($n = 3755$). Com base na variável de identificação animal, presente em todos os arquivos e única a cada animal, foram descartadas observações incompletas entre os arquivos ($n = 1901$) e, por fim, todas as informações foram agrupadas em um único arquivo ($n = 3267$).

Seis variáveis adicionais foram criadas com base nas informações disponíveis. Com base na data do nascimento das vacas e a data do primeiro parto desses animais, foram criadas as variáveis estação do nascimento (primavera, verão, outono e inverno) e estação do parto (primavera, verão, outono e inverno), respectivamente. Foram ainda estimados o mês do nascimento e o mês do primeiro parto tendo como base as mesmas variáveis. A relação gordura/proteína foi calculada com base nas produções desses constituintes do leite (kg dia^{-1}) obtidas no primeiro controle leiteiro. Foi criada ainda variável indicando a ocorrência ou não de parto gemelar.

Parcelas perdidas encontradas nas variáveis nitrogênio uréico do leite ($n = 710$), facilidade de parto ($n = 184$), sexo ($n = 97$), tamanho ($n = 237$) e mortalidade ($n = 7$) da cria foram estimadas utilizando a técnica de *K-nearest neighbours* estabelecendo $k = 5$ por meio da função *KNN* do pacote *VIM* (KOWARIK; TEMPL, 2016).

2.2 Seleção de variáveis

Variáveis não relacionadas com o primeiro controle leiteiro foram descartadas, sendo, portanto mantidas 27 variáveis que posteriormente foram avaliadas quanto a sua capacidade preditiva.

Variáveis relacionadas com a P_LEITE foram identificadas utilizando a *regularized regression* (regressão regularizada) por meio da *elastic net* (rede elástica), a qual combina as penalidades das regressões regularizadas *Least Absolute Shrinkage and Selection Operator* (LASSO) e *Ridge* (ZOU; HASTIE, 2005). Previamente, o banco de dados foi normalizado para média 0 e desvio padrão 1. Os parâmetros α , responsável por controlar o número de variáveis selecionadas, e λ , responsável pela elasticidade da rede, foram estimados por meio de *10-fold cross validation* testando 100 possíveis valores variando entre 0 e 1, utilizando a função *cv.glmnet* do pacote *glmnet* (FRIEDMA; HASTIE; TIBSHIRANI, 2010). Em seguida, a função *train* do pacote *caret* (KUHNS, 2018) foi utilizada para gerar o modelo final, especificando os valores de α e λ previamente estimados. Para o presente estudo, foram encontrados os valores de 0,46 e 0,52, respectivamente para α e λ , selecionando 15 variáveis.

O banco de dados efetivamente utilizado nas análises posteriores foi então, constituído da variável reposta P_LEITE, juntamente com as variáveis selecionadas. Este foi dividido entre arquivo de treino (n = 2452) e validação (n = 815) seguindo a proporção 75:25 (Tabela 1 Tabela 2), respectivamente, utilizando a função *initial_split* do pacote *rsample* (KUHNS; WICKHAM, 2017). O arquivo de treino foi utilizado para a estimativa dos parâmetros dos diferentes modelos avaliados ao passo que o arquivo de validação foi utilizado para avaliar a performance dos modelos.

Tabela 1 – Estatísticas descritivas da variável dependente (produção de leite no primeiro controle) e variáveis quantitativas, selecionadas pela *regularized regression* via *elastic net*, nos arquivos de treino e validação.

Variável	Treino ⁽¹⁾	Validação ⁽¹⁾
	(n = 2452)	(n = 815)
Produção de leite no primeiro controle (kg)	29,5 ± 6,03	29,4 ± 6,09
	29,6	29,5
	10,0 – 53,2	11,5 – 55,6
Dias em lactação ⁽²⁾	21,7 ± 9,77	21,7 ± 9,72
	22,0	21,0
	5 – 40	5 – 40
Nitrogênio uréico no leite (mg dL ⁻¹) ⁽³⁾	10,6 ± 3,04	10,7 ± 2,87
	10,5	10,6
	0,0 – 36,5	2,6 – 19,8
Contagem de células somáticas (10 ³ células mL ⁻¹) ⁽³⁾	159,9 ± 518,13	148,9 ± 439,78
	49,0	46,0
	1,0 – 9999,0	3,0 – 6379,0
Score linear da contagem de células somáticas ⁽³⁾	2,3 ± 1,60	2,2 ± 1,56
	2,0	1,9
	0,1 – 9,6	0,1 – 9,0
β-hidroxibutirato no leite (mmol L ⁻¹) ⁽³⁾	0,10 ± 0,06	0,10 ± 0,07
	0,09	0,09
	0,0 – 0,77	0,0 – 1,10
Relação gordura/proteína ⁽³⁾	1,33 ± 0,27	1,32 ± 0,27
	1,30	1,28
	0,33 – 3,96	0,51 – 3,33
Peso corporal ao parto (Kg)	615,9 ± 63,57	618,3 ± 61,12
	614,8	614,2
	454,8 – 988,0	436,1 – 898,3
Ganho médio diário entre 12 e 15 meses de idade (Kg/animal/dia)	0,750 ± 0,09	0,760 ± 0,09
	0,750	0,750
	0,520 – 1,110	0,530 – 1,000
Mês do parto ⁽⁴⁾	7,0 ± 3,56	7,2 ± 3,49
	7,3	7,4
	1,0 – 12,9	1,0 – 12,9

⁽¹⁾ Média ± desvio padrão; mediana; amplitude. ⁽²⁾ Dias em lactação no primeiro controle leiteiro. ⁽³⁾ Mensurado no dia do primeiro controle leiteiro. ⁽⁴⁾ Leva em consideração o mês e o dia de ocorrência do parto, variando entre 1,0 e 12,9, indicando a ocorrência do parto no primeiro dia de janeiro e no último dia de dezembro, respectivamente.

Tabela 2 – Distribuição das variáveis qualitativas, selecionadas pela *regularized regression* via *elastic net*, nos arquivos de treino e validação.

Variável	Treino (n = 2452)		Validação (n = 815)	
	N	%	N	%
Coleta da amostra de leite ⁽¹⁾				
24 horas	1505	61,38	504	61,84
Manhã	460	18,76	144	17,67
Tarde	444	18,11	156	19,14
Outro	43	1,75	11	1,35
Frequência de ordenha ⁽²⁾				
1	321	13,09	105	12,88
2	2057	83,89	698	85,64
3	74	3,02	12	1,47
Início da lactação				
Parto normal	2426	98,94	808	99,14
Aborto	26	1,06	7	0,86
Mortalidade da cria				
Sobreviveu	2155	87,89	727	89,2
Não sobreviveu	297	12,11	88	10,8
Ocorrência de parto gemelar				
Sim	14	0,57	3	0,37
Não	2438	99,43	812	99,63
Estação do parto ⁽³⁾				
Primavera	543	22,15	181	22,21
Verão	664	27,08	218	26,75
Outono	571	23,29	206	25,28
Inverno	674	27,49	210	25,77

⁽¹⁾ Variável categórica indicando a ordenha utilizada para a coleta de amostra de leite pelos técnicos de campo da *Valacta*. Posteriormente, foi utilizada também pela *Valacta* para ajustar a produção de leite, assim como seus constituintes, no primeiro controle leiteiro de acordo com diretrizes estabelecidas pelo *International Committee for Registration of Animals* (ICAR, 2017). ⁽²⁾ Número de ordenhas por dia. ⁽³⁾ Primavera = 20 de março a 20 de junho, Verão = 21 de junho a 21 de setembro, Outono = 22 de setembro a 20 de dezembro, Inverno = 21 de dezembro a 19 de março.

2.3 Modelagens avaliadas

No presente trabalho foram testadas três modelagens distintas: regressão linear multivariada, *random forest*, rede neural artificial.

2.3.1 Regressão linear multivariada

O arquivo de treino foi submetido à regressão linear multivariada (RLM) utilizando a função *lm* do pacote *stats*, adotando o seguinte modelo de regressão linear múltipla:

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_n + \varepsilon \quad (1)$$

em que Y é a variável dependente, X_1, X_2, \dots, X_n são as variáveis independentes, $\alpha_1, \alpha_2, \dots, \alpha_n$ são os coeficientes de regressão, e ε é o erro residual. Nesse estudo, P_LEITE foi considerada a variável dependente e as variáveis selecionadas pela *regularized regression via elastic net* foram as variáveis independentes.

A normalidade dos resíduos foi avaliada pelo teste de Shapiro-Wilk (SHAPIRO; WILK, 1965) conjuntamente com o gráfico QQ-Normal. A homoscedasticidade foi avaliada pelo teste de Breusch-Pagan (BREUSCH; PAGAN, 1979) e pelo gráfico dos resíduos padronizados versus valores preditos. Ainda, a independência do resíduo foi avaliada por meio do teste de Durbin-Watson (DURBIN; WATSON, 1950, 1951).

Para o modelo e testes utilizados, foi adotado nível de significância de 5% de probabilidade.

2.3.2 Random forest

Random forest (RF) consiste na combinação de árvores de decisão $r_X(x, \beta_k)$, $k = 1, 2, \dots, n$, onde β_k é o subconjunto de variáveis aleatoriamente selecionadas e utilizadas como critério de partição recursiva das árvores individuais (BREIMAN, 2001a). Dessa forma, os parâmetros internos (hiperparâmetros), número e árvores (n_{tree}) e β_k devem ser definidos. O valor ótimo de β_k foi determinado em $\beta_k = 5$, utilizando *10-fold cross validation* através da função *train* do pacote *caret* (KUHN, 2018) definindo *rf* como método (LIAW; WIENER, 2002). O n_{tree} deve ser grande o suficiente para permitir que cada observação seja predita mais de uma vez durante o treinamento do modelo, sendo então definido $n_{tree} = 2000$ no presente trabalho.

2.3.3 Rede neural artificial

Convencionalmente, uma rede neural artificial (RNA) é composta por unidades processadoras simples (neurônios) arranjadas em camadas interconectadas. Os neurônios da camada de entrada são ativados pelas variáveis preditivas apresentadas ao modelo. Essa informação é então transformada pela aplicação de uma função de ativação e transmitida para as demais camadas (camadas ocultas). Os demais neurônios são então ativados de acordo com o peso das ligações com os neurônios de camadas anteriores, sucessivamente, até a camada de saída, onde o modelo retorna o resultado (SCHMIDHUBER, 2015). Quanto maior o número de

camadas ocultas, maior é a capacidade da RNA de se ajustar a distribuições não-lineares de maior complexidade (SIRIGNANO; SADHWANI; GIESECKE, 2018), sendo capazes de aproximar qualquer distribuição, dada quantidade suficiente de neurônios (HORNIK, 1991).

Antes de serem submetidas à RNA, variáveis qualitativas dos arquivos de treino e validação (Tabela 2) foram *one-hot encoded*, utilizando a função *dummyVars* do pacote *caret* (KUHNS, 2018). Em seguida, tanto variáveis quantitativas (Tabela 1) quanto qualitativas (Tabela 2) foram normalizadas para terem média 0 e desvio padrão 1.

Nesse estudo, *grid search* foi utilizado na definição do número de camadas ocultas, quantidade de neurônios em cada camada oculta e a função de ativação. Foram treinadas 288 redes diferentes, resultado da combinação única entre os possíveis valores avaliados para o número de camadas ocultas (1, 2, 3 ou 4), número de neurônios em cada camada oculta (50, 100, ou 200) e função de ativação nas camadas de entrada e ocultas [*exponential linear unit* (Equação 2; CLEVERT et al., 2015) ou tangente hiperbólica (Equação 3)]:

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

$$f(x) = \begin{cases} x & x > 0 \\ \alpha (e^x - 1) & x \leq 0 \end{cases}, \quad (3)$$

em que x , em ambas as equações acima, representa o somatório ponderado das entradas. Visando reduzir a ocorrência de *overfitting* e garantir a capacidade de generalização da RNA, também foram avaliados modelos com fatores de regularização L1 (0 ou 1×10^{-3}) e L2 (0 ou 1×10^{-3}), além da taxa de *dropout* [0, 25 ou 50% (SRIVASTAVA et al., 2014)].

Apenas redes simétricas foram avaliadas, ou seja, redes nas quais o número de neurônios em todas as camadas ocultas era o mesmo, uma vez que modelos com essa arquitetura têm apresentado melhores resultados quando comparados a diferentes arquiteturas (LAROUELLE et al., 2009).

Não foi necessário avaliar diferentes taxas de aprendizado, pois foi utilizado o algoritmo ADADELTA na otimização dos pesos, no qual a taxa de aprendizado se adapta dinamicamente durante o treinamento da RNA (ZEILER, 2012).

A função de ativação linear:

$$f(x) = x, \quad (4)$$

em que x representa a somatória dos pesos, foi utilizada em todos os modelos entre a última camada oculta e a camada de saída. Além disso, todos os modelos possuíam 26 neurônios na camada de entrada, representando as variáveis selecionadas pela *regularized regression* via *elastic net*, e 1 neurônio na camada de saída, representando L_LEITE (Tabela 1 e Tabela 2).

Os modelos foram construídos utilizando o arquivo de treino definindo *epoch*¹ máximo de 500 e *batch*² com tamanho de 20 observações. Durante o treinamento, foi adotado critério de *early stopping*, também visando maximizar a capacidade de generalização da RNA, além de potencialmente reduzir o tempo de treinamento (COULIBALY; ANCTIL; BOBÉE, 2000). Por esse critério, o treinamento prosseguiu por *n epochs* até que a melhora na acurácia de predição do modelo, mensurada no conjunto de validação ao fim de cada *epoch*, não fosse maior que 1×10^{-4} por 20 *epochs* consecutivos. O erro absoluto médio (Tabela 3) foi a estatística utilizada para mensurar a acurácia dos modelos.

Os pesos (*W*) das ligações entre os neurônios foram inicializados de acordo com Glorot e Bengio (2010) (Equação 5):

$$W \sim U \left[-\frac{\sqrt{6}}{\sqrt{n_i + n_j}}, \frac{\sqrt{6}}{\sqrt{n_i + n_j}} \right] \quad (5)$$

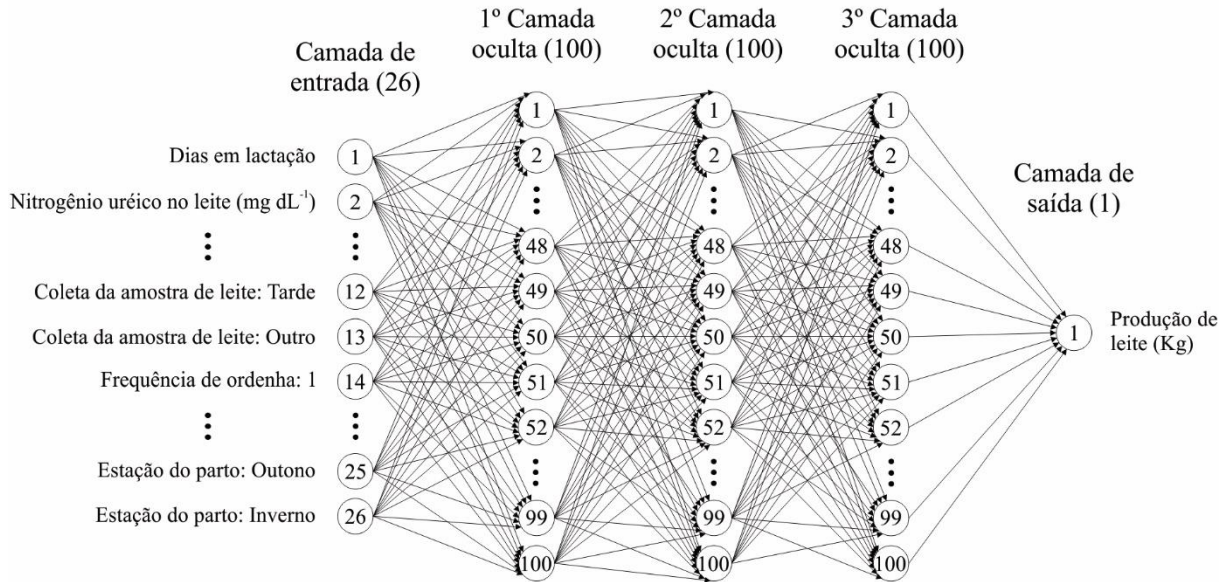
onde n_i é a quantidade de neurônios de entrada e n_j é a quantidade de neurônios de saída. Em todos os modelos da *grid search*, foi aplicada *batch normalization* com supressão do termo *bias*. Por esta metodologia, os pesos das ligações são normalizados para terem média 0 e desvio padrão 1 a cada *batch* de treinamento, antes de serem passados para a função de ativação, de modo a reduzir o tempo necessário para o treinamento da RNA até a convergência e maximiza a performance dos modelos (IOFFE; SZEGEDY, 2015).

O treinamento das RNAs e seleção da melhor arquitetura foi realizado utilizando a interface do Keras (ALLAIRE; CHOLLET, 2018), implementada no programa R (R CORE TEAM, 2018). O melhor modelo encontrado através da *grid search* possuía três camadas ocultas com 100 neurônios em cada (Figura 6), fator de regularização $L1 = 1 \times 10^{-3}$ e função de ativação *exponential linear unit* (CLEVERT; UNTERTHINER; HOCHREITER, 2015) (Equação 2), sendo esse utilizado nas análises posteriores.

¹A cada *epoch*, o modelo foi apresentado a todo o conjunto de dados que foi utilizado no treinamento, sendo esse um processo iterativo que se repete por *n* vezes até que critério de convergência seja atendido.

²Número de observações a serem utilizadas em cada iteração de treinamento do modelo

Figura 6 – Rede neural artificial utilizada no presente estudo contendo 26 neurônios na camada de entrada, 100 neurônios em cada uma das três camadas ocultas e um neurônio na camada de saída. Neurônios são interligados entre as camadas e a força das ligações é determinada através dos pesos (w) associados a elas. A informação transmitida por um neurônio aos neurônios da camada seguinte é resultado da somatória dos pesos que chegam até ele, transformada pela aplicação de uma função de ativação (g).



2.4 Avaliação dos modelos

Seis critérios foram utilizados na avaliação da performance dos modelos. A eficiência do ajuste foi avaliada através do coeficiente de determinação (R^2), ao passo que, o desvio entre os valores observados e preditos foram avaliados pela raiz do erro médio quadrático (REMQ), erro absoluto médio (EAM), erro percentual médio (EPM), erro percentual médio absoluto simétrico (EPMAS) e a razão entre os desvios padrões (DPr) dos valores preditos e observados (Tabela 3). Foi considerado como melhor modelo aquele com o maior valor de R^2 e menores valores de REMQ, EAM, EPM, EPAMS e DPr.

Tabela 3 – Critérios de avaliação do ajuste dos modelos e as respectivas fórmulas utilizadas.

Critério	Formula ⁽¹⁾
Coeficiente de determinação	$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2}$
Raiz do erro médio quadrático	$REMQ = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}$
Erro absoluto médio	$EAM = \frac{\sum_{i=1}^n Y_i - \hat{Y}_i }{n}$
Erro percentual médio	$EPM = \frac{\sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}}}{\bar{Y}} \times 100$
Erro percentual médio absoluto simétrico	$EPMAS = \frac{100\%}{n} \sum_{i=1}^n \frac{ Y_i - \hat{Y}_i }{(Y_i + \hat{Y}_i)/2}$
Razão do desvio padrão	$DPr = \frac{DP_{pred}}{DP_{obs}}$

⁽¹⁾ Y_i = Produção de leite observada no primeiro controle leiteiro no i^{th} animal; \hat{Y}_i = Produção de leite predita para o primeiro controle leiteiro no i^{th} animal; \bar{Y} = Média observada de produção de leite no primeiro controle leiteiro; n = Número de observações; DP_{pred} = Desvio padrão da produção de leite predita para o primeiro controle; DP_{obs} = Desvio padrão observado para a produção de leite no primeiro controle leiteiro.

Além disso, *10-fold cross validation* também foi aplicada no conjunto de dados completo (treino e validação), a fim de avaliar o ajuste dos três modelos. O conjunto de dados foi aleatoriamente dividido em 10 partes independentes, sendo nove partes utilizadas no treinamento do modelo e uma parte utilizada para avaliação dos resultados (KOHAVI, 1995). Esse procedimento foi realizado $k = 10$ vezes.

A capacidade preditiva dos modelos foi avaliada da seguinte forma: uma vez que o teste de Shapiro-Wilk (SHAPIRO; WILK, 1965) indicou que tanto a P_LEITE observada quanto a estimada pelos modelos, não apresentavam distribuição normal de probabilidade ($P < 0,05$), foi utilizado o teste não paramétrico de Wilcoxon (WILCOXON, 1945), com ajuste para medidas pareadas para comparação entre os valores preditos e observados. Além disso, foram calculados o coeficiente de correlação linear de Pearson (r) e o coeficiente de concordância de Lin (LIN, 1989).

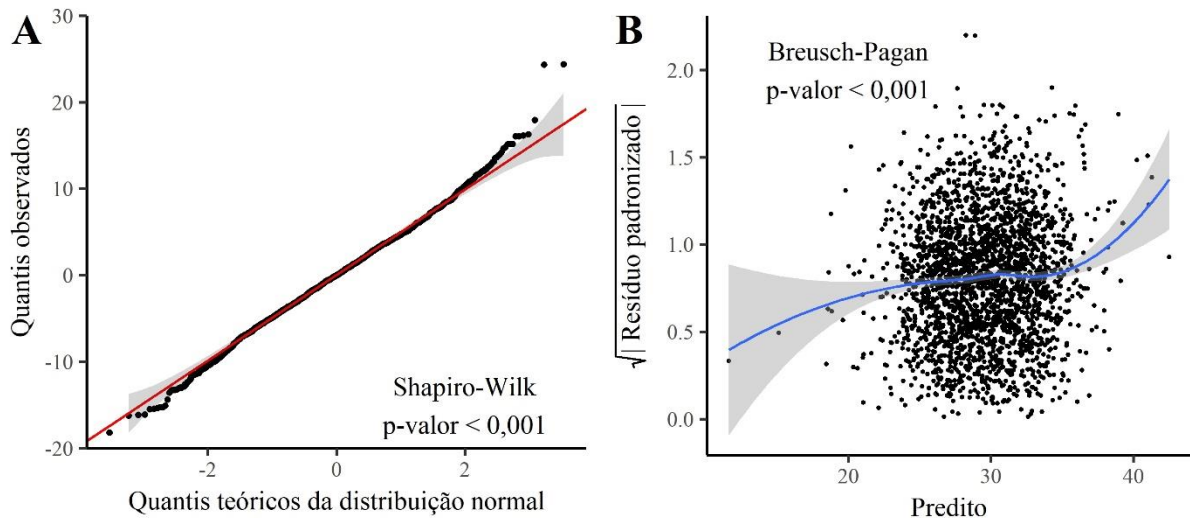
A fim de entender como os diferentes modelos integraram as variáveis preditivas, objetivando prever P_LEITE, foi utilizada a metodologia *Local Interpretable Model-Agnostic Explanations* (LIME) proposta por Ribeiro, Singh e Guestrin (2016) e implementada no R através do pacote *lime* (PEDERSEN; BENESTY, 2018), nos três modelos avaliados no presente trabalho (RLM, RF e RNA). A função *lime* foi utilizada para criar explicações baseadas no

arquivo de treino e posteriormente, utilizadas na função *explain* para criar explicações para o arquivo de validação, com base nas variáveis mais importantes de acordo com a regressão *Ridge*.

3 RESULTADOS

A análise dos resíduos da regressão linear multivariada mostrou o não atendimento das pressuposições inerentes à metodologia. A homoscedasticidade e a normalidade do resíduo (Figura 7), assim como a independência do resíduo, pressuposições necessárias a um modelo RLM, foram todas violadas ($P < 0,05$) de acordo com os testes utilizados, tornando questionável a real efetividade deste modelo para o objetivo proposto.

Figura 7 – (A) Gráfico Quantil-Quantil dos resíduos da regressão linear com linha de tendência e intervalo de 95% de probabilidade (cinza), juntamente com o resultado do teste de Shapiro-Wilk indicando a violação da pressuposição de normalidade ($P < 0,001$). (B) Valores preditos pela regressão linear versus resíduos padronizados com linha de tendência e intervalo de 95% de probabilidade (cinza), juntamente com o resultado do teste de Breusch-Pagan indicando a violação da pressuposição de homoscedasticidade ($P < 0,001$).



A RNA apresentou os melhores resultados de ajuste, sendo ligeiramente superiores à RLM e RF, muito embora todos tenham apresentado qualidade de ajuste semelhante (Tabela 4). A RLM e a RNA não apresentaram grandes variações entre as avaliações realizadas no arquivo de treino, de validação e pela *cross validation*, indício de que não ocorreu *overfitting* nestes modelos. Por outro lado, foi encontrada grande diferença entre as estatísticas de ajustes obtidas para a RF, entre o arquivo de treino e validação, indicando que ocorreu *overfitting*, reduzindo assim, a capacidade de generalização desse modelo.

Tabela 4 – Qualidade do ajuste dos modelos de regressão linear multivariada, *random forest* e rede neural artificial obtido para o arquivo de treino (n = 2452), validação (n = 815) e *10-fold cross validation* seguido pelo intervalo de confiança com 95% de probabilidade (IC 95%).

Item	Regressão linear multivariada			<i>Random forest</i>			Rede neural artificial		
	Treino	Validação	<i>Cross validation</i> ⁽¹⁾ (IC 95%)	Treino	Validação	<i>Cross validation</i> ⁽¹⁾ (IC 95%)	Treino	Validação	<i>Cross validation</i> ⁽¹⁾ (IC 95%)
R ² ⁽²⁾	0,30	0,30	0,29 (0,27 – 0,32)	0,93	0,32	0,30 (0,28 – 0,33)	0,34	0,32	0,32 (0,28 – 0,36)
REMQ ⁽³⁾	5,04	5,10	5,08 (4,89 – 5,28)	0,25	5,07	5,05 (4,90 – 5,21)	4,91	5,02	4,98 (4,77 – 5,19)
EAM ⁽⁴⁾	3,94	3,99	3,98 (3,88 – 4,09)	1,73	3,94	3,94 (3,87 – 4,02)	3,80	3,89	3,88 (3,75 – 4,00)
EPM ⁽⁵⁾	17,07	17,33	17,23 (16,60 – 17,87)	7,64	17,21	17,14 (16,63 – 17,65)	16,64	17,05	16,89 (16,32 – 17,46)
EPMAS ⁽⁶⁾	13,64	13,99	13,84 (13,46 – 14,22)	6,17	13,77	13,71 (13,41 – 14,01)	13,16	13,62	13,48 (13,15 – 13,81)
DPr ⁽⁷⁾	0,55	0,52	0,55 (0,53 – 0,57)	0,70	0,48	0,50 (0,49 – 0,52)	0,54	0,54	0,54 (0,51 – 0,58)

⁽¹⁾ *10-fold cross validation*. ⁽²⁾ Coeficiente de determinação. ⁽³⁾ Raiz do erro médio quadrático. ⁽⁴⁾ Erro absoluto médio. ⁽⁵⁾ Erro percentual médio. ⁽⁶⁾ Erro percentual médio absoluto simétrico. ⁽⁷⁾ Razão do desvio padrão.

O coeficiente de correlação de Pearson entre os valores observados e preditos para os arquivos de treino e validação apresentaram comportamento similar às estatísticas de ajuste (Tabela 5). No arquivo de treino, o maior valor foi observado para RF seguido pela RNA e RLM. Por outro lado, no arquivo de validação o maior valor foi observado para RNA, RF e RLM.

Tabela 5 – Coeficiente de correlação de Pearson (r) e coeficiente de concordância de Lin (ρ) seguido pelo intervalo de confiança com 95% de probabilidade (IC 95%) obtidos entre os valores preditos e observados para o arquivo de treino (n = 2452) e validação (n = 815).

Modelo	Treino			Validação		
	r	p -valor	ρ (IC 95%)	r	p -valor	ρ (IC 95%)
Regressão linear multivariada	0,55	< 0,001	0,46 (0,44 – 0,49)	0,55	< 0,001	0,45 (0,40 – 0,49)
<i>Random forest</i>	0,96	< 0,001	0,91 (0,90 – 0,91)	0,56	< 0,001	0,44 (0,40 – 0,48)
Rede neural artificial	0,58	< 0,001	0,49 (0,46 – 0,51)	0,57	< 0,001	0,48 (0,43 – 0,52)

Similar ao coeficiente de correlação de Pearson, o coeficiente de concordância de Lin [ρ (LIN, 1989)] varia entre 1 e -1, representando respectivamente, completa concordância e discordância, ao passo que 0 indica a independência das observações (CARRASCO *et al.*, 2013). De acordo com ρ no arquivo de treino, o melhor modelo foi a RF, não sendo o mesmo observado no arquivo de validação, onde o melhor modelo foi a RNA. Os resultados apresentados na Tabela 5 também apresentam indício de *overfitting* na RF.

A comparação entre os valores medianos preditos de P_LEITE para o arquivo de validação indicou que, tanto as predições feitas pela RLM quanto pela RF e RNA, não foram estatisticamente diferentes das observadas ($P > 0,05$), sugerindo que, em valores medianos, a capacidade preditiva dos modelos foi similar (Tabela 6).

Tabela 6 – Comparação entre a produção de leite mediana no primeiro controle leiteiro observada e predita pelos modelos de regressão linear multivariada, *random forest* e rede neural artificial no arquivo de validação (n = 815).

Modelo	Produção de leite		P - valor
	Observada	Predita	
	(Mediana \pm desvio absoluto)	(Mediana \pm desvio absoluto)	
Regressão linear multivariada	29,44 \pm 5,93	29,53 \pm 3,40	0,61
<i>Random forest</i>	29,44 \pm 5,93	30,09 \pm 3,03	0,24
Rede neural artificial	29,44 \pm 5,93	29,58 \pm 3,39	0,37

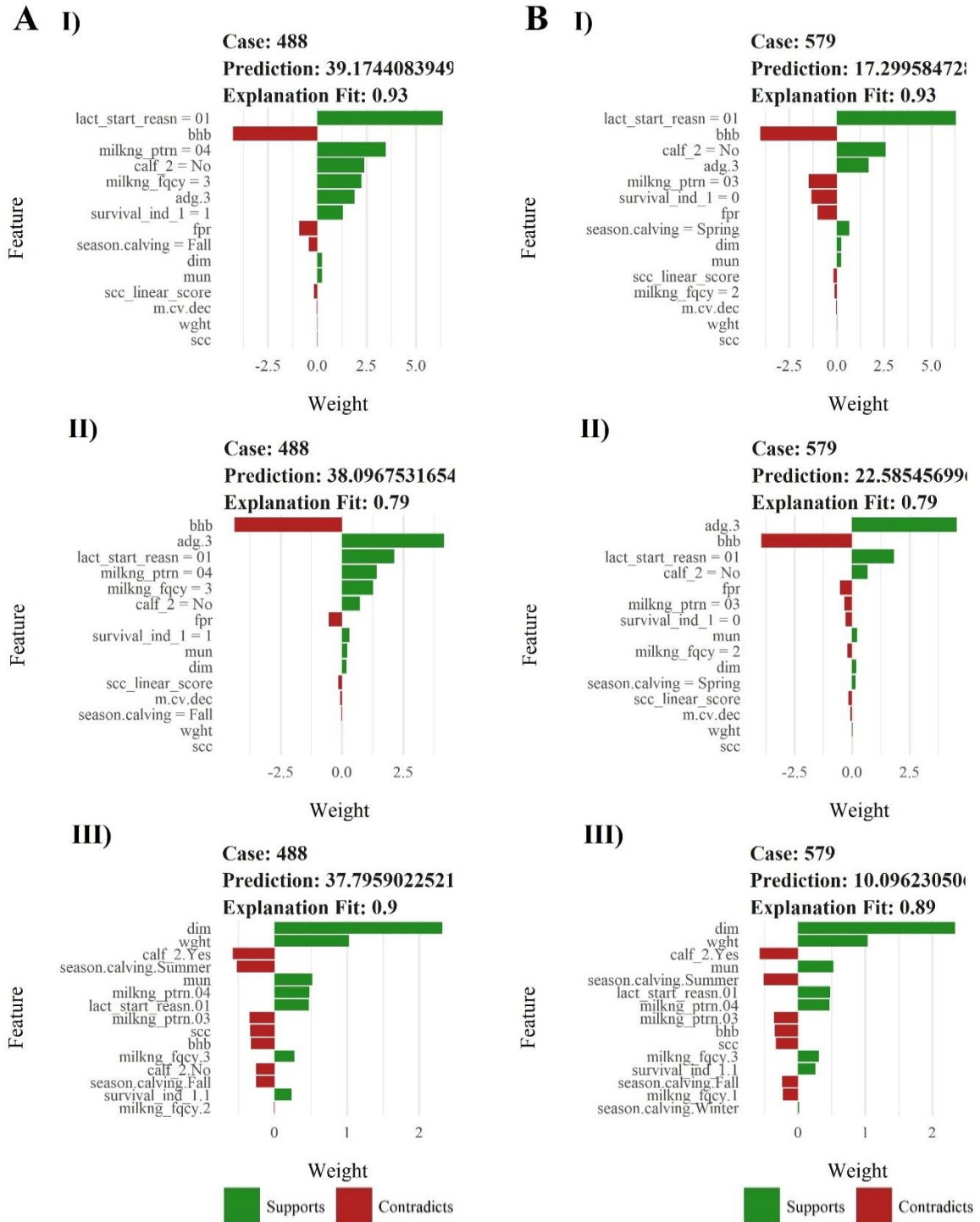
A interpretação dos modelos segundo a metodologia LIME proposta por Ribeiro, Singh e Guestrin (2016) assume que a relação entre as variáveis preditivas e variável resposta

seja localmente linear, podendo ser modelada por modelo linear esparso, gerando a importância relativa de cada variável, especificamente para cada predição, juntamente com o coeficiente de determinação (R^2), que indica a qualidade da avaliação (RIBEIRO; SINGH; GUESTRIN, 2016). Os coeficientes de determinação (média \pm desvio padrão) obtidos para as avaliações dos modelos RLM, RF e RNA foram $0,92 \pm 0,015$, $0,78 \pm 0,006$ e $0,89 \pm 0,004$, respectivamente, indicando boa qualidade da avaliação. Foi possível observar que as variáveis que mais influenciaram nas predições foram relativamente similares entre RLM e RM, ao passo que RNA diferiu dos demais modelos (Figura 8).

As variáveis que mais influenciaram as predições da RLM foram início da lactação com parto normal e o teor de β -hidroxibutirato no leite, sendo então observado efeito sinérgico e antagônico dessas variáveis, tanto nas predições maiores quanto menores, respectivamente [Figura 8 – A (I) e B(I)].

A variável teor de β -hidroxibutirato no leite também foi a que mais influenciou negativamente as predições da RF, ao passo que o ganho médio diário de peso entre 12 e 15 meses de idade foi a variável com maior peso favorável à P_LEITE, seguida pela variável que indica que a lactação teve início a partir de um parto normal [Figura 8. A (II) e B (II)]. Por outro lado, dias em lactação no dia do teste e peso ao parto, foram as variáveis com maior peso favorável à predição de P_LEITE, ao passo que parto gemelar foi a variável com maior peso contrário à predição de P_LEITE pela RNA [Figura 8 – A (III) e B (III)].

Figura 8 – Gráficos da importância relativa das variáveis nas maiores (A) e menores (B) predições nos modelos de regressão linear multivariada (I), *random forest* (II) e rede neural artificial (III).



Legenda: lact_start_reasn = 1: início da lactação com parto normal; bbb: β-hidroxibutirato no leite; milkng_ptrn = 03: coleta de amostra código 24 horas; milkng_ptrn = 04: coleta de amostra código outro; calf_2 = No: Ausência de parto gemelar; calf_2 = Yes: Ocorreu parto gemelar; milkng_fqcy = 1: Frequência de ordenha de uma vez ao dia; milkng_fqcy = 2: Frequência de ordenha de duas vezes ao dia; milkng_fqcy = 3: Frequência de ordenha de três vezes ao dia; adg.3: ganho médio diário entre 12 e 15 meses de idade; survival_ind_1 = 0: Cria morreu; survival_ind_1 = 1: Cria não morreu; fpr: relação gordura proteína do leite; season.calving = Fall: lactação teve início com parto no outono; season.calving = Summer: lactação teve início com parto no verão; season.calving = Winter: lactação teve início com parto no inverno; dim: dias em lactação no dia do controle leiteiro; mun: teor de ureia no leite; scc: contagem de células somáticas no leite; scc_linear_score: score linear da contagem de células somáticas do leite; m.cv.dec: mês do parto; wght: peso ao parto.

4 DISCUSSÃO

O propósito no uso das técnicas de regressão linear múltipla (RLM), *random forest* (RF) e rede neural artificial (RNA) no presente estudo foi o de prever a produção de leite no primeiro controle de vacas primíparas (P_LEITE), buscando assim, desenvolver uma ferramenta passível de ser utilizada na avaliação do período de transição de animais nessa categoria, de forma análoga ao utilizado para vacas múltíparas (NORDLUND, 2006).

A P_LEITE tem sido apresentada como representativa do sucesso ou não do período de transição. Heuer, Schukken e Dobbelaar (1999) encontraram que a produção de leite, juntamente com a relação gordura/proteína (FPD), melhor indicou a ocorrência de problemas durante a transição, do que a variação no peso corporal entre o pré e pós-parto. Além disso, a maior incidência de distúrbios metabólicos observada durante o parto (ØSTERGAARD; GRÖHN, 1999) afeta negativamente a P_LEITE (GANTNER; BOBIĆ; POTOČNIK, 2016), o que se estende por toda a lactação (DRACKLEY, 1999). Dessa forma, a produção de leite obtida durante o período de transição reflete o resultado do manejo adotado, justificando assim o uso da P_LEITE como forma de avaliá-lo.

Os três modelos avaliados apresentaram capacidade preditiva para a P_LEITE no presente trabalho (Tabela 6). À luz do conhecimento, nenhum outro trabalho prévio a esse foi conduzido com o objetivo de prever a P_LEITE, impossibilitando avaliar a performance relativa dos modelos aqui testados com os de outros autores, muito embora a comparação entre modelos descritos em diferentes trabalhos é dificultada pelo fato de serem caso-específicos e utilizarem variáveis distintas na maioria das vezes (MURPHY *et al.*, 2014). Por outro lado, o uso de RNA se mostrou bem sucedido quando empregado na modelagem de dados da pecuária leiteira, visando prever a produção de leite em 305 dias de lactação (GRZESIAK; BŁASZCZYK; LACROIX, 2006; MURPHY *et al.*, 2014; NJUBI; WAKHUNGU; BADAMANA, 2010; SHARMA; SHARMA; KASANA, 2007), assim como prever o início do trabalho de parto (BORCHERS *et al.*, 2017) e a necessidade de auxílio durante o parto (FENLON *et al.*, 2017). Já a RF tem sido utilizada na análise de informações genéticas, como a identificação de polimorfismo entre nucleotídeos associados com o consumo alimentar de vacas leiteiras (YAO *et al.*, 2013).

Os resultados observados para RF no presente trabalho indicaram que ocorreu *overfitting* desse modelo ao arquivo de treino, reduzindo assim sua capacidade de generalização, como observado nos resultados obtidos no arquivo de validação e *10-fold cross validation*, comparado com os resultados no arquivo de treino (Tabela 4 e Tabela 5). *Overfitting* refere-se à reduzida capacidade de realizar previsões corretas, por modelos que adaptaram

perfeitamente às amostras de treino, uma vez que apreendem não somente a tendência estrutural dos dados, mas também o ruído presente (STROBL; MALLEY; TUTZ, 2009). Breiman (2001b) sugeriu que RF não sofre de *overfitting* ao avaliar a performance desse tipo de modelo em alguns bancos de dados *benchmarking* do repositório da *University of California, Irvine* (UCI). Entretanto, Segal (2004) avaliando os mesmos bancos de dados, concluiu que modelos criados com base em árvores de decisão, dificilmente sofreriam de *overfitting* naqueles arquivos, podendo o mesmo não ser verdade em outros arquivos. Os resultados aqui encontrados (Tabela 4 e Tabela 5) indicam que esse modelo sofre de *overfitting*, muito embora não se saiba ainda o real motivo por trás da ocorrência desse efeito (STROBL; MALLEY; TUTZ, 2009).

O modelo que apresentou os melhores resultados de ajuste foi a RNA, muito embora a melhora na performance não tenha sido de grande magnitude, quando comparado aos demais modelos testados (Tabela 4). Por outro lado, RNA apresenta maior flexibilidade quando comparado à RLM, já que esse tipo de modelo não assume condições de homoscedasticidade e normalidade dos resíduos (ADAMCZYK *et al.*, 2016), e tais condições não foram observadas no presente trabalho (Figura 7).

O maior peso antagônico à P_LEITE nos modelos RLM e RF foi atribuído ao β -hidroxibutirato no leite (Figura 8). O teor de β -hidroxibutirato encontrado no leite pode ser utilizado como indicador da prevalência de cetose nos animais (DENIS-ROBICHAUD *et al.*, 2014), sendo esse um dos distúrbios metabólicos que mais acomete o gado leiteiro durante o período de transição, com taxa de incidência variando entre 26 e 56% (MCART; NYDAM; OETZEL, 2012). Especificamente para primíparas, Santschi *et al.* (2016) encontraram que, em média, 18,7% dos animais eram acometidos, sendo a maior prevalência (31%) observada logo na primeira semana pós-parto. Além disso, a cetose é considerada o principal fator de risco para outros distúrbios e doenças (GRÖHN *et al.*, 1989; SHIN *et al.*, 2015; SUTHAR *et al.*, 2013), podendo reduzir em até 4,3 vezes a chance da vaca emprenhar na primeira inseminação (RUTHERFORD; OIKONOMOU; SMITH, 2016). Primíparas acometidas por esse distúrbio podem apresentar redução na P_LEITE variando entre 8% e 10%, quando comparadas a animais saudáveis (BENEDET *et al.*, 2017; SANTSCHI *et al.*, 2016). Dessa forma, é biologicamente justificável a relação negativa entre a variável β -hidroxibutirato no leite e a P_LEITE, tornando plausível o efeito atribuído a ela pela RLM e RF.

As variáveis que mais influenciaram positivamente a predição da P_LEITE na RNA foram o peso ao parto, assim como, dias em lactação no dia do controle leiteiro (Figura 8). Com base na curva de lactação, observa-se aumento diário na produção de leite ao longo das primeiras semanas pós-parto até o pico de lactação, sendo então observada redução gradual na

produção de leite (WOOD, 1967). Dessa forma, seria esperado e desejável observar relação positiva entre dias em lactação e P_LEITE, muito embora o mesmo não tenha sido observado nos modelos RLM e RF. Além disso, peso ao parto tem sido relacionado positivamente com a performance produtiva de novilhas leiteiras (MACDONALD *et al.*, 2005; STASSEN *et al.*, 1991), justificando biologicamente a relação positiva entre essa informação e a P_LEITE, assim como observada nas predições da RNA.

Por outro lado, a variável indicando a ocorrência ou não de parto gemelar foi a que apresentou maior peso negativo em relação às predições da P_LEITE pela RNA (Figura 8). Embora a prevalência de partos gêmeares tenha sido baixa no presente trabalho (Tabela 2), esse tipo de parto tem sido associado com aumento significativo no risco dos animais desenvolverem problemas reprodutivos como metrite e retenção de placenta, assim como cetose (MARKUSFELD, 1987; NIELEN *et al.*, 1989) e partos distócicos (HOSSEIN-ZADEH, 2010). Consequentemente, a P_LEITE tende a ser menor (GHAVI HOSSEIN-ZADEH, 2013), sendo sugerido a interrupção de gestações gêmeares, como alternativa para a redução do seu impacto econômico negativo na atividade leiteira (MUR-NOVALES *et al.*, 2018).

Trabalhos futuros poderiam avaliar a possível variação entre rebanhos, de forma a identificar tendências (BROTZMAN *et al.*, 2015; DALLAGO *et al.*, 2017) e a sua respectiva inclusão nos modelos. O fator rebanho se mostrou importante na avaliação da resposta imune de vacas em transição (ZECCONI *et al.*, 2018), assim como na avaliação da incidência de distúrbios metabólicos, desempenho reprodutivo e taxa de descarte de animais (HEUER; SCHUKKEN; DOBBELAAR, 1999), e como fator de risco para aborto (KESHAVARZI *et al.*, 2017). Dessa forma, seria razoável levantar a hipótese de que a inclusão desse fator na modelagem poderia melhorar a capacidade preditiva dos modelos, principalmente da RNA, a julgar pelo efeito negativo desses problemas na P_LEITE.

5 CONCLUSÃO

Os três modelos avaliados apresentaram capacidade preditiva para a P_LEITE. RNA foi o modelo que apresentou a melhor qualidade de ajuste, sendo demonstrada a possibilidade de uso desse tipo de modelagem, em dados oriundos da pecuária leiteira. Trabalhos futuros poderão avaliar variações entre rebanhos, como forma de melhorar a capacidade preditiva dos modelos.

REFERÊNCIAS

ADAMCZYK, K. *et al.* Recognition of culling reasons in Polish dairy cows using data mining methods. **Computers and Electronics in Agriculture**, v. 127, p. 26-37, 2016.

ALLAIRE, J. J.; CHOLLET, F. **keras: R Interface to 'Keras'**. <https://CRAN.R-project.org/package=keras>. R Package Versão 2.1.5, 2018.

BARLETTA, R. V. *et al.* Association of changes among body condition score during the transition period with NEFA and BHBA concentrations, milk production, fertility, and health of Holstein cows. **Theriogenology**, v. 104, p. 30-36, 2017.

BARNES, M. A.; PEARSON, R. E.; LUKES-WILSON, A. J. Effects of milking frequency and selection for milk yield on productive efficiency of holstein cows. **Journal of Dairy Science**, v. 73, n. 6, p. 1603-1611, 1990.

BAUMAN, D. E.; CURRIE, W. B. Partitioning of nutrients during pregnancy and lactation: A review of mechanisms involving homeostasis and homeorhesis. **Journal of Dairy Science**, v. 63, n. 9, p. 1514-1529, 1980.

BEEVER, D. E. The impact of controlled nutrition during the dry period on dairy cow health, fertility and performance. **Animal Reproduction Science**, v. 96, n. 3, p. 212-226, 2006.

BELL, A. W. Regulation of organic nutrient metabolism during transition from late pregnancy to early lactation. **Journal of Animal Science**, v. 73, n. 9, p. 2804-2819, 1995.

BENEDET, A. *et al.* Effects of ketosis status defined by FTIR spectroscopy on milk quality traits of first-lactation cows. **Agriculturae Conspectus Scientificus**, v. 82, n. 2, p. 167-170, 2017.

BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. In: MONTAVON, G.; ORRKLAUS, G. B.; MÜLLER, R. **Neural Networks: Tricks of the Trade**. 2 ed. Berlin, Heidelberg: Springer, 2012. cap. 19, p. 437 - 478.

BORCHERS, M. R. *et al.* Machine-learning-based calving prediction from activity, lying, and ruminating behaviors in dairy cattle. **Journal of Dairy Science**, v. 100, n. 7, p. 5664-5674, 2017.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, n. 1, p. 5-32, 2001a.

BREIMAN, L. Statistical modeling: The two cultures. **Statistical Science**, v. 16, n. 3, p. 199-231, 2001b.

BREUSCH, T. S.; PAGAN, A. R. A simple test for heteroscedasticity and random coefficient variation. **Econometrica**, v. 47, n. 5, p. 1287-1294, 1979.

BROTZMAN, R. L. *et al.* Cluster analysis of Dairy Herd Improvement data to discover trends in performance characteristics in large Upper Midwest dairy herds. **Journal of Dairy Science**, v. 98, n. 5, p. 3059-3070, 2015.

CARRASCO, J. L. *et al.* Estimation of the concordance correlation coefficient for repeated measures using SAS and R. **Computer Methods and Programs in Biomedicine**, v. 109, n. 3, p. 293-304, 2013.

CHAGAS, L. M. *et al.* Invited Review: New Perspectives on the Roles of Nutrition and Metabolic Priorities in the Subfertility of High-Producing Dairy Cows. **Journal of Dairy Science**, v. 90, n. 9, p. 4022-4032, 2007.

CHAPINAL, N. *et al.* The association of serum metabolites in the transition period with milk production and early-lactation reproductive performance. **Journal of Dairy Science**, v. 95, n. 3, p. 1301-1309, 2012.

CLEVERT, D. A.; UNTERTHINER, T.; HOCHREITER, S. Fast and accurate deep network learning by exponential linear units (elus). **arXiv preprint arXiv:1511.07289**, 2015.

COMPTON, C. W. R.; YOUNG, L.; MCDUGALL, S. Subclinical ketosis in post-partum dairy cows fed a predominantly pasture-based diet: defining cut-points for diagnosis using concentrations of beta-hydroxybutyrate in blood and determining prevalence. **New Zealand Veterinary Journal**, v. 63, n. 5, p. 241-248, 2015.

COULIBALY, P.; ANCTIL, F.; BOBÉE, B. Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. **Journal of Hydrology**, v. 230, n. 3, p. 244-257, 2000.

CUE, R. I. *et al.* Growth modeling of dairy heifers in Québec based on random regression. **Canadian Journal of Animal Science**, v. 92, n. 1, p. 33-47, 2012.

DALLAGO, G. M. *et al.* Análise de cluster dos rebanhos leiteiros do estado de Minas Gerais. In: CAMPOS, P. H. R. F. *et al.*, Simpósio Mineiro de Produção Animal, 4. e Semana da Zootecnia, 7., 2017, Diamantina, Minas Gerais, Brasil. **Anais...** Diamantina, Minas Gerais, Brasil: UFVJM, 2017.

DAROS, R. R. *et al.* Prevalence and risk factors for transition period diseases in grazing dairy cows in Brazil. **Preventive Veterinary Medicine**, v. 145, p. 16-22, 2017.

DENIS-ROBICHAUD, J. *et al.* Accuracy of milk ketone bodies from flow-injection analysis for the diagnosis of hyperketonemia in dairy cows. **Journal of Dairy Science**, v. 97, n. 6, p. 3364-3370, 2014.

DOUGLAS, G. N. *et al.* Prepartal plane of nutrition, regardless of dietary energy source, affects periparturient metabolism and dry matter intake in Holstein cows. **Journal of Dairy Science**, v. 89, n. 6, p. 2141-2157, 2006.

DRACKLEY, J. K. Biology of dairy cows during the transition period: The final frontier? **Journal of Dairy Science**, v. 82, n. 11, p. 2259-2273, 1999.

DRACKLEY, J. K.; CARDOSO, F. C. Prepartum and postpartum nutritional management to optimize fertility in high-yielding dairy cows in confined TMR systems. **Animal**, v. 8, n. s1, p. 5-14, 2014.

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression I. **Biometrika**, v. 37, n. 3-4, p. 409-428, 1950.

DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression II. **Biometrika**, v. 38, n. 1-2, p. 159-178, 1951.

EDWARDS, J. L.; TOZER, P. R. Using activity and milk yield as predictors of fresh cow disorders. **Journal of Dairy Science**, v. 87, n. 2, p. 524-531, 2004.

FENLON, C. *et al.* A comparison of 4 predictive models of calving assistance and difficulty in dairy heifers and cows. **Journal of Dairy Science**, v. 100, n. 12, p. 9746-9758, 2017.

FIORE, E. *et al.* Adaptation of some energetic parameters during transition period in dairy cows. **Journal of Applied Animal Research**, v. 46, n. 1, p. 402-405, 2018.

FRIEDMA, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. **Journal of Statistical Software**, v. 33, n. 1, p. 1-22, 2010.

GANTNER, V.; BOBIĆ, T.; POTOČNIK, K. Prevalence of metabolic disorders and effect on subsequent daily milk quantity and quality in Holstein cows. **Archives Animal Breeding**, v. 59, n. 3, p. 381-386, 2016.

GARVERICK, H. A. *et al.* Concentrations of nonesterified fatty acids and glucose in blood of periparturient dairy cows are indicative of pregnancy success at first insemination. **Journal of Dairy Science**, v. 96, n. 1, p. 181-188, 2013.

GEISHAUSER, T. *et al.* Fat/protein ratio in first DHI test milk as test for displaced abomasum in dairy cows. **Journal of Veterinary Medicine Series A**, v. 44, n. 1-10, p. 265-270, 1997.

GHAVI HOSSEIN-ZADEH, N. Effects of main reproductive and health problems on the performance of dairy cows: a review. **Spanish Journal of Agricultural Research**, v. 11, n. 3, p. 18, 2013.

GHEISE, N. J. E. *et al.* Effect of pre-calving body condition score and previous lactation on BCS change, blood metabolites, oxidative stress and milk production in Holstein dairy cows. **Italian Journal of Animal Science**, v. 16, n. 3, p. 474-483, 2017.

GILLUND, P. *et al.* Body condition related to ketosis and reproductive performance in Norwegian dairy cows. **Journal of Dairy Science**, v. 84, n. 6, p. 1390-1396, 2001.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: YEE WHYE, T.; MIKE, T., Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, 2010, Proceedings of Machine Learning Research. **Anais...** PMLR, 2010.

GOFF, J. P. The monitoring, prevention, and treatment of milk fever and subclinical hypocalcemia in dairy cows. **The Veterinary Journal**, v. 176, n. 1, p. 50-57, 2008.

GOFF, J. P.; HORST, R. L. Physiological changes at parturition and their relationship to metabolic disorders. **Journal of Dairy Science**, v. 80, n. 7, p. 1260-1268, 1997.

GONZÁLEZ, L. A. *et al.* Changes in feeding behavior as possible indicators for the automatic monitoring of health disorders in dairy cows. **Journal of Dairy Science**, v. 91, n. 3, p. 1017-1028, 2008.

GRÖHN, Y. T. *et al.* Epidemiology of metabolic disorders in dairy cattle: Association among host characteristics, disease, and production. **Journal of Dairy Science**, v. 72, n. 7, p. 1876-1885, 1989.

GRUMMER, R. R. Impact of changes in organic nutrient metabolism on feeding the transition dairy cow. **Journal of Animal Science**, v. 73, n. 9, p. 2820-2833, 1995.

GRUMMER, R. R.; MASHEK, D. G.; HAYIRLI, A. Dry matter intake and energy balance in the transition period. **Veterinary Clinics of North America: Food Animal Practice**, v. 20, n. 3, p. 447-470, 2004.

GRZESIAK, W.; BŁASZCZYK, P.; LACROIX, R. Methods of predicting milk yield in dairy cows - Predictive capabilities of Wood's lactation curve and artificial neural networks (ANNs). **Computers and Electronics in Agriculture**, v. 54, n. 2, p. 69-83, 2006.

HAMMON, D. S. *et al.* Neutrophil function and energy status in Holstein cows with uterine health disorders. **Veterinary Immunology and Immunopathology**, v. 113, n. 1, p. 21-29, 2006.

HEUER, C.; SCHUKKEN, Y. H.; DOBBELAAR, P. Postpartum body condition score and results from the first test day milk as predictors of disease, fertility, yield, and culling in commercial dairy herds. **Journal of Dairy Science**, v. 82, n. 2, p. 295-304, 1999.

HORNIK, K. Approximation capabilities of multilayer feedforward networks. **Neural Networks**, v. 4, n. 2, p. 251-257, 1991.

HOSSEIN-ZADEH, N. G. The effect of twinning on milk yield, dystocia, calf birth weight and open days in Holstein dairy cows of Iran. **Journal of Animal Physiology and Animal Nutrition**, v. 94, n. 6, p. 780-787, 2010.

HUZZEY, J. M. *et al.* Prepartum behavior and dry matter intake identify dairy cows at risk for metritis. **Journal of Dairy Science**, v. 90, n. 7, p. 3220-3233, 2007.

ICAR. Procedure 1 of section 2 of ICAR guidelines - Computing 24-hour yields. **The global standard for livestock data**. Roma, Itália: International Committee for Registration of Animals - ICAR, 2017. p.17. Disponível em: < <https://www.icar.org/Guidelines/02-Procedure-1-Computing-24-Hour-Yield.pdf> >. Acesso em: 28 de junho de 2018.

INCHAISRI, C. *et al.* Economic consequences of reproductive performance in dairy cattle. **Theriogenology**, v. 74, n. 5, p. 835-846, 2010.

IOFFE, S.; SZEGEDY, C. **Batch normalization: Accelerating deep network training by reducing internal covariate shift**. 2015. ArXiv e-prints. arXiv:1502.03167v3. Disponível em: < <https://arxiv.org/abs/1502.03167v3> >. Acesso em: 23 de junho de 2018.

JORDAN, E. R.; FOURDRAINE, R. H. Characterization of the Management Practices of the Top Milk Producing Herds in the Country. **Journal of Dairy Science**, v. 76, n. 10, p. 3247-3256, 1993.

KEHRLI, M. E.; NONNECKE, B. J.; ROTH, J. A. Alterations in bovine neutrophil function during the periparturient period. **American Journal of Veterinary Research**, v. 50, n. 2, p. 207-214, 1989.

KESHAVARZI, H. *et al.* Abortion studies in Iranian dairy herds: I. Risk factors for abortion. **Livestock Science**, v. 195, p. 45-52, 2017.

KIMURA, K.; REINHARDT, T. A.; GOFF, J. P. Parturition and hypocalcemia blunts calcium signals in immune cells of dairy cattle. **Journal of Dairy Science**, v. 89, n. 7, p. 2588-2595, 2006.

KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: MELLISH, C. S., International Joint Conference on Artificial Intelligence, 14., 1995, Montreal, Canadá. **Anais...** San Francisco, Califórnia, EUA: Morgan Kaufmann, 1995.

KOWARIK, A.; TEMPL, M. Imputation with the R package VIM. **Journal of Statistical Software**, v. 74, n. 7, p. 1-16, 2016.

KUHN, M. **caret: classification and regression training**. <https://CRAN.R-project.org/package=caret>. R Package Versão 6.0-79, 2018.

KUHN, M.; WICKHAM, H. **rsample: General resampling infrastructure**. <https://CRAN.R-project.org/package=rsample>. R Package Versão 0.0.2, 2017.

LACASSE, P. *et al.* Innovative dairy cow management to improve resistance to metabolic and infectious diseases during the transition period. **Research in Veterinary Science**, v. 116, p. 40-46, 2018.

LACETERA, N. *et al.* Short communication: Effects of nonesterified fatty acids on lymphocyte function in dairy heifers. **Journal of Dairy Science**, v. 87, n. 4, p. 1012-1014, 2004.

LAROCHELLE, H. *et al.* Exploring strategies for training deep neural networks. **Journal of Machine Learning Research**, v. 1, p. 1-40, 2009.

LEBLANC, S. Monitoring metabolic health of dairy cattle in the transition period. **Journal of Reproduction and Development**, v. 56 Suppl, p. S29-35, 2010.

LEBLANC, S. J. *et al.* Major advances in disease prevention in dairy cattle. **Journal of Dairy Science**, v. 89, n. 4, p. 1267-1279, 2006.

- LEYS, C. *et al.* Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. **Journal of Experimental Social Psychology**, v. 49, n. 4, p. 764-766, 2013.
- LIAW, A.; WIENER, M. Classification and regression by randomForest. **R News**, v. 2, n. 3, p. 18-22, 2002.
- LIN, L. I. K. A concordance correlation coefficient to evaluate reproducibility. **Biometrics**, v. 45, n. 1, p. 255-268, 1989.
- LOISELLE, M. C. *et al.* Impact of postpartum milking frequency on the immune system and the blood metabolite concentration of dairy cows. **Journal of Dairy Science**, v. 92, n. 5, p. 1900-1912, 2009.
- LUKAS, J. M. *et al.* A study of methods for evaluating the success of the transition period in early-lactation dairy cows. **Journal of Dairy Science**, v. 98, n. 1, p. 250-262, 2015.
- MACDONALD, K. A. *et al.* Effect of feeding level pPre- and post-puberty and body weight at first calving on growth, milk production, and fertility in grazing dairy cows. **Journal of Dairy Science**, v. 88, n. 9, p. 3363-3375, 2005.
- MANN, S. *et al.* Dry period plane of energy: Effects on feed intake, energy balance, milk production, and composition in transition dairy cows. **Journal of Dairy Science**, v. 98, n. 5, p. 3366-3382, 2015.
- MARKUSFELD, O. Periparturient traits in seven high dairy herds. Incidence rates, association with parity, and interrelationships among traits. **Journal of Dairy Science**, v. 70, n. 1, p. 158-166, 1987.
- MARTINEZ, N. *et al.* Evaluation of peripartal calcium status, energetic profile, and neutrophil function in dairy cows at low or high risk of developing uterine disease. **Journal of Dairy Science**, v. 95, n. 12, p. 7158-7172, 2012.
- MCART, J. A. A.; NYDAM, D. V.; OETZEL, G. R. Epidemiology of subclinical ketosis in early lactation dairy cattle. **Journal of Dairy Science**, v. 95, n. 9, p. 5056-5066, 2012.
- MCART, J. A. A. *et al.* A field trial on the effect of propylene glycol on milk yield and resolution of ketosis in fresh cows diagnosed with subclinical ketosis. **Journal of Dairy Science**, v. 94, n. 12, p. 6011-6020, 2011.
- MCGRATH, B. A. *et al.* Composition and properties of bovine colostrum: a review. **Dairy Science & Technology**, v. 96, n. 2, p. 133-158, 2016.
- MUR-NOVALES, R. *et al.* An economic evaluation of management strategies to mitigate the negative effect of twinning in dairy herds. **Journal of Dairy Science**, v. 101, n. 9, p. 1-15, 2018.
- MURPHY, M. D. *et al.* Comparison of modelling techniques for milk-production forecasting. **Journal of Dairy Science**, v. 97, n. 6, p. 3352-3363, 2014.

NEAVE, H. W. *et al.* Parity differences in the behavior of transition dairy cows. **Journal of Dairy Science**, v. 100, n. 1, p. 548-561, 2017.

NIELEN, M. *et al.* Twinning in dairy cattle: A study of risk factors and effects. **Theriogenology**, v. 32, n. 5, p. 845-862, 1989.

NJUBI, D. M.; WAKHUNGU, J. W.; BADAMANA, M. S. Use of test-day records to predict first lactation 305-day milk yield using artificial neural network in Kenyan Holstein–Friesian dairy cows. **Tropical Animal Health and Production**, v. 42, n. 4, p. 639–644, 2010.

NORDLUND, K. V. Transition Cow Index™. American Association Bovine Practitioners, 39., 2006, St. Paul, MN. **Anais...** Stillwater, OK: Frontier Printers, 2006.

NORDLUND, K. V. *et al.* **Method for optimizing health and productivity of milk producing animals**. US 7886691 B2, 30/março/2006, 15/fevereiro/2011.

NORDLUND, K. V.; COOK, N. B. Using herd records to monitor transition cow survival, productivity, and health. **Veterinary Clinics of North America: Food Animal Practice**, v. 20, n. 3, p. 627-649, 2004.

NRC. **Nutrient requirements of dairy cattle**. 7th. Washington, DC: The National Academies Press, 2001. 408 p. ISBN 978-0-309-06997-7.

OETZEL, G. R. Monitoring and testing dairy herds for metabolic disease. **Veterinary Clinics of North America: Food Animal Practice**, v. 20, n. 3, p. 651-674, 2004.

ØSTERGAARD, S.; GRÖHN, Y. T. Effects of diseases on test may milk yield and body weight of dairy cows from Danish research herds. **Journal of Dairy Science**, v. 82, n. 6, p. 1188-1201, 1999.

PALMQUIST, D. L.; MATTOS, W. R. S. Metabolismo de lipídios. In: BERCHIELLI, T. T.; PIREZ, A. V.; OLIVEIRA, S. G. D. **Nutrição de ruminantes**. 2 ed. Jaboticabal: Funep, 2011. p. 299-321.

PEDERSEN, T. L.; BENESTY, M. **lime: Local Interpretable Model-Agnostic Explanations**. <https://CRAN.R-project.org/package=lime>. R Package Versão 0.4.0, 2018.

PENNER, G. B.; BEAUCHEMIN, K. A.; MUTSVANGWA, T. Severity of ruminal acidosis in primiparous holstein cows during the periparturient period. **Journal of Dairy Science**, v. 90, n. 1, p. 365-375, 2007.

PINEDA, A.; CARDOSO, F. C. Technical note: Validation of a handheld meter for measuring β -hydroxybutyrate concentrations in plasma and serum from dairy cows. **Journal of Dairy Science**, v. 98, n. 12, p. 8818-8824, 2015.

PINEDO, P. J.; DE VRIES, A.; WEBB, D. W. Dynamics of culling risk with disposal codes reported by Dairy Herd Improvement dairy herds. **Journal of Dairy Science**, v. 93, n. 5, p. 2250-2261, 2010.

PONCHEKI, J. K. **Avaliação do manejo de vacas no período de transição utilizando as informações do primeiro controle leiteiro após o parto.** 2015. 90 f. Dissertação (Masters) – Departamento de Medicina Veterinária, Universidade Federal do Paraná, Curitiba, Paraná, 2015.

PROUDFOOT, K. L.; HUZZEY, J. M.; VON KEYSERLINGK, M. A. G. The effect of dystocia on the dry matter intake and behavior of Holstein cows. **Journal of Dairy Science**, v. 92, n. 10, p. 4937-4944, 2009.

R CORE TEAM. **R: A language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing. Versão 3.4.4 "Someone to Lean On", 2018.

RAJALA-SCHULTZ, P. J.; GRÖHN, Y. T.; MCCULLOCH, C. E. Effects of Milk Fever, Ketosis, and Lameness on Milk Yield in Dairy Cows. **Journal of Dairy Science**, v. 82, n. 2, p. 288-294, 1999.

REINHARDT, T. A. *et al.* Prevalence of subclinical hypocalcemia in dairy herds. **The Veterinary Journal**, v. 188, n. 1, p. 122-124, 2011.

RENSIS, F. D.; SCARAMUZZI, R. J. Heat stress and seasonal effects on reproduction in the dairy cow - A review. **Theriogenology**, v. 60, n. 6, p. 1139-1151, 2003.

REYNOLDS, C. K. *et al.* Splanchnic metabolism of dairy cows during the transition from late gestation through early lactation. **Journal of Dairy Science**, v. 86, n. 4, p. 1201-1217, 2003.

RIBEIRO, E. S. *et al.* Prevalence of periparturient diseases and effects on fertility of seasonally calving grazing dairy cows supplemented with concentrates. **Journal of Dairy Science**, v. 96, n. 9, p. 5682-5697, 2013.

RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. **"Why should I trust you?": Explaining the predictions of any classifier.** 2016. [ArXiv e-prints](https://arxiv.org/abs/1602.04938v3). arXiv:1602.04938v3. Disponível em: <<https://arxiv.org/abs/1602.04938v3>>. Acesso em: 27 de junho de 2018.

ROBERTS, T. *et al.* Metabolic parameters in transition cows as indicators for early-lactation culling risk. **Journal of Dairy Science**, v. 95, n. 6, p. 3057-3063, 2012.

ROCHE, J. R. *et al.* Nutritional management of the transition cow in the 21st century – A paradigm shift in thinking. **Animal Production Science**, v. 53, n. 9, p. 1000-1023, 2013.

RODRÍGUEZ, E. M.; ARÍS, A.; BACH, A. Associations between subclinical hypocalcemia and postparturient diseases in dairy cows. **Journal of Dairy Science**, v. 100, n. 9, p. 7427-7434, 2017.

RUTHERFORD, A. J.; OIKONOMOU, G.; SMITH, R. F. The effect of subclinical ketosis on activity at estrus and reproductive performance in dairy cattle. **Journal of Dairy Science**, v. 99, n. 6, p. 4808-4815, 2016.

SANTOS, J. E. P. Distúrbios metabólicos. In: BERCHIELLI, T. T.; PIREZ, A. V.; OLIVEIRA, S. G. D. **Nutrição de ruminantes.** 2 ed. ed. Jaboticabal: FUNEP, 2011. p. 439-520.

SANTSCHI, D. E. *et al.* Prevalence of elevated milk β -hydroxybutyrate concentrations in Holstein cows measured by Fourier-transform infrared analysis in Dairy Herd Improvement milk samples and association with milk yield and components. **Journal of Dairy Science**, v. 99, n. 11, p. 9263-9270, 2016.

SCALIA, D. *et al.* In vitro effects of nonesterified fatty acids on bovine neutrophils oxidative burst and viability. **Journal of Dairy Science**, v. 89, n. 1, p. 147-154, 2006.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. **Neural Networks**, v. 61, p. 85-117, 2015.

SCHULTZ, K. K. *et al.* Exploring relationships between Dairy Herd Improvement monitors of performance and the Transition Cow Index in Wisconsin dairy herds. **Journal of Dairy Science**, v. 99, n. 9, p. 7506-16, 2016.

SEGAL, M. R. **Machine learning benchmarks and random forest regression**. 2004. Disponível em: < <https://escholarship.org/uc/item/35x3v9t4> >. Acesso em: 05 de agosto de 2018.

SEIFI, H. A.; KIA, S. Subclinical hypocalcemia in dairy cows: pathophysiology, consequences and monitoring. **Iranian Journal of Veterinary Science and Technology**, v. 9, n. 2, p. 1-15, 2017.

SEIFI, H. A. *et al.* Metabolic predictors of post-partum disease and culling risk in dairy cattle. **The Veterinary Journal**, v. 188, n. 2, p. 216-220, 2011.

SEPÚLVEDA-VARAS, P. *et al.* Transition diseases in grazing dairy cows are related to serum cholesterol and other analytes. **PloS One**, v. 10, n. 3, p. e0122317, 2015.

SHAPIRO, S. S.; WILK, M. B. An analysis of variance test for normality (complete samples). **Biometrika**, v. 52, n. 3-4, p. 591-611, 1965.

SHARMA, A. K.; SHARMA, R. K.; KASANA, H. S. Prediction of first lactation 305-day milk yield in Karan Fries dairy cattle using ANN modeling. **Applied Soft Computing**, v. 7, n. 3, p. 1112-1120, 2007.

SHIN, E. K. *et al.* Relationships among ketosis, serum metabolites, body condition, and reproductive outcomes in dairy cows. **Theriogenology**, v. 84, n. 2, p. 252-260, 2015.

SHRESTHA, S. **Energy status and reproductive performances in dairy cattle in New Zealand**. 2015. 74 f. Dissertação (Mestrado em Estudos Veterinários) – Massey University, Palmerston North, Nova Zelândia, 2015.

SIRIGNANO, J. A.; SADHWANI, A.; GIESECKE, K. **Deep learning for mortgage risk**. 2018. Disponível em: < <https://dx.doi.org/10.2139/ssrn.2799443> >. Acesso em: 26 de junho de 2018.

SRIVASTAVA, N. *et al.* Dropout: A simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, v. 15, p. 1929-1958, 2014.

STASSEN, E. N. *et al.* Age and weight at first calving and subsequent performance in Dutch dairy cattle from rearing farms. **Veterinary Quarterly**, v. 13, n. 3, p. 180-184, 1991.

STER, C.; LOISELLE, M. C.; LACASSE, P. Effect of postcalving serum nonesterified fatty acids concentration on the functionality of bovine immune cells. **Journal of Dairy Science**, v. 95, n. 2, p. 708-717, 2012.

STROBL, C.; MALLEY, J.; TUTZ, G. An introduction to recursive partitioning: Rationale, application and characteristics of classification and regression trees, bagging and random forests. **Psychological Methods**, v. 14, n. 4, p. 323-348, 2009.

SUTHAR, V. S. *et al.* Prevalence of subclinical ketosis and relationships with postpartum diseases in European dairy cows. **Journal of Dairy Science**, v. 96, n. 5, p. 2925-2938, 2013.

TURK, R. *et al.* Lipid mobilisation and oxidative stress as metabolic adaptation processes in dairy heifers during transition period. **Animal Reproduction Science**, v. 141, n. 3, p. 109-115, 2013.

VACEK, M. *et al.* Relationships between growth and body condition development during the rearing period and performance in the first three lactations in Holstein cows. **Czech Journal of Animal Science**, v. 60, n. 9, p. 417-425, 2015.

VAN DORP, R. T. *et al.* An epidemiologic study of disease in 32 registered Holstein dairy herds in British Columbia. **Canadian Journal of Veterinary Research**, v. 63, n. 3, p. 185-192, 1999.

VANHOLDER, T. *et al.* Risk factors for subclinical and clinical ketosis and association with production parameters in dairy cows in the Netherlands. **Journal of Dairy Science**, v. 98, n. 2, p. 880-888, 2015.

WEAVER, S. R. *et al.* Serotonin and calcium homeostasis during the transition period. **Domestic Animal Endocrinology**, v. 56, p. S147-S154, 2016.

WEST, J. W. Effects of heat-stress on production in dairy cattle. **Journal of Dairy Science**, v. 86, n. 6, p. 2131-2144, 2003.

WILCOXON, F. Individual comparisons by ranking methods. **Biometrics Bulletin**, v. 1, n. 6, p. 80-83, 1945.

WOOD, P. D. P. Algebraic model of the lactation curve in cattle. **Nature**, v. 216, p. 164, 1967.

YAO, C. *et al.* Random forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. **Journal of Dairy Science**, v. 96, n. 10, p. 6716-6729, 2013.

ZEBELI, Q. *et al.* Nutrition, rumen health and inflammation in the transition period and their role on overall health and fertility in dairy cows. **Research in Veterinary Science**, v. 103, p. 126-136, 2015.

ZECCONI, A. *et al.* Effects of herd and physiological status on variation of 16 immunological and inflammatory parameters in dairy cows during drying off and the transition period. **Journal of Dairy Research**, v. 85, n. 2, p. 167-173, 2018.

ZEILER, M. D. **ADADELTA: An adaptive learning rate method**. 2012. ArXiv e-prints. arXiv:1212.5701. Disponível em: < <https://arxiv.org/abs/1212.5701> >. Acesso em: 23 de junho de 2018.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 67, n. 2, p. 301-320, 2005.

