

UNIVERSIDADE DE SÃO PAULO  
FACULDADE DE ZOOTECNIA E ENGENHARIA DE ALIMENTOS

ANDRÉ LEVI VIANA PEREIRA

**Modelo baseado em aprendizado de máquina para classificar o conforto  
térmico de bezerras usando termografia por infravermelho**

---

Pirassununga

2022

ANDRÉ LEVI VIANA PEREIRA

**Modelo baseado em aprendizado de máquina para classificar o conforto  
térmico de bezerras usando termografia por infravermelho**

Versão corrigida

Dissertação apresentada à Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo, como parte dos requisitos para a obtenção do título de Mestre em Ciências do programa de Mestrado em Zootecnia

Área de Concentração: Qualidade e Produtividade Animal

Orientador: Prof. Dr. Rafael Vieira de Sousa

---

Pirassununga

2022

Ficha catalográfica elaborada pelo  
Serviço de Biblioteca e Informação, FZEA/USP,  
com os dados fornecidos pelo(a) autor(a)

P436m      Pereira, André Levi Viana  
Modelo baseado em aprendizado de máquina para classificar o conforto térmico de bezerras usando termografia por infravermelho / André Levi Viana Pereira ; orientador Rafael Vieira de Sousa. -- Pirassununga, 2022.  
73 f.

Dissertação (Mestrado - Programa de Pós-Graduação em Zootecnia) -- Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo.

1. Bem-Estar Animal. 2. Estresse Térmico. 3. Mineração de Dados. 4. Modelagem Computacional. 5. Termografia por Infravermelho. I. Sousa, Rafael Vieira de, orient. II. Título.



## CERTIFICADO

Certificamos que a proposta intitulada "Respostas termorregulatórias de bezerras holandesas submetidas a ondas de calor: avaliação do estresse agudo e crônico", protocolada sob o CEUA nº 6957201219 (ID 001415), sob a responsabilidade de **Luciane Silva Martello** - que envolve a produção, manutenção e/ou utilização de animais pertencentes ao filo Chordata, subfilo Vertebrata (exceto o homem), para fins de pesquisa científica ou ensino - está de acordo com os preceitos da Lei 11.794 de 8 de outubro de 2008, com o Decreto 6.899 de 15 de julho de 2009, bem como com as normas editadas pelo Conselho Nacional de Controle da Experimentação Animal (CONCEA), e foi **aprovada** pela Comissão de Ética no Uso de Animais da Faculdade de Zootecnia e Engenharia de Alimentos da Universidade de São Paulo - FZEA/USP (CEUA/FZEA) na reunião de 28/01/2020.

We certify that the proposal "Thermoregulatory responses of Holstein calves submitted to heat waves: evaluation of acute and chronic stress", utilizing 10 Bovines (males and females), protocol number CEUA 6957201219 (ID 001415), under the responsibility of **Luciane Silva Martello** - which involves the production, maintenance and/or use of animals belonging to the phylum Chordata, subphylum Vertebrata (except human beings), for scientific research purposes or teaching - is in accordance with Law 11.794 of October 8, 2008, Decree 6899 of July 15, 2009, as well as with the rules issued by the National Council for Control of Animal Experimentation (CONCEA), and was **approved** by the Ethic Committee on Animal Use of the School of Animal Science and Food Engineering - (São Paulo University) (CEUA/FZEA) in the meeting of 01/28/2020.

Finalidade da Proposta: [Pesquisa \(Acadêmica\)](#)

Vigência da Proposta: de [05/2020](#) a [07/2020](#)

Área: [Engenharia de Biosistemas](#)

Origem: [Faculdade de Zootecnia e Engenharia de Alimentos \(FZEA/USP\)](#)

Espécie: [Bovinos](#)

sexo: [Machos e Fêmeas](#)

idade: [60 a 70 dias](#)

N: [10](#)

Linhagem: [Raça Holandesa](#)

Peso: [120 a 140 kg](#)

Local do experimento: Câmara climática do Departamento de Reprodução Animal, pertencente à Faculdade de Medicina Veterinária e Zootecnia da Universidade de São Paulo (VRA/FMVZ) e baía coletiva, anexa à câmara climática.

Pirassununga, 28 de janeiro de 2020

Profa. Dra. Daniele dos Santos Martins  
Coordenadora da Comissão de Ética no Uso de Animais  
Faculdade de Zootecnia e Engenharia de Alimentos da  
Universidade de São Paulo - FZEA/USP

Profa. Dra. Cristiane Gonçalves Titto  
Vice-Coordenadora da Comissão de Ética no Uso de Animais  
Faculdade de Zootecnia e Engenharia de Alimentos da  
Universidade de São Paulo - FZEA/USP

## DEDICATÓRIA

Dedico este trabalho à minha mãe, que tanto se esforçou para me educar da melhor maneira possível. Sem ela eu não seria nada do que sou hoje.

## **AGRADECIMENTOS**

Agradeço a todos os professores que participaram da minha formação e que contribuíram para meu conhecimento e capacidade para realizar este projeto.

Agradeço à Universidade de São Paulo e à Faculdade de Zootecnia e Engenharia de Alimentos pela estrutura de ensino e pesquisa oferecida, bem como ao pagador de impostos do estado de São Paulo por garantir o funcionamento dessa estrutura.

Agradeço aos colegas de laboratório Jéssica, Alex, Isabele, Maria e Gabriel por estarem presentes durante a execução deste projeto e pela amizade construída.

Agradeço ao técnico de laboratório Edson, pela ajuda prestada para a realização deste projeto.

Agradeço à Professora Luciane, que contribuiu com a minha coorientação.

Agradeço ao Professor Rafael, que além de ser um excelente orientador ainda esteve presente nos momentos de dificuldade.

Por fim, agradeço à minha mãe, minha maior professora e a quem eu devo tudo o que sou.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## RESUMO

PEREIRA, A. L. V. **Modelo baseado em aprendizado de máquina para classificar o conforto térmico de bezerras usando termografia por infravermelho.** 2022. 74 f. Dissertação (Mestrado) – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2022.

Pesquisas recentemente realizadas no âmbito da produção animal relacionam o estresse térmico dos animais a efeitos negativos sobre sua saúde e produtividade. Estresse térmico é uma condição que prejudica o desempenho produtivo e reprodutivo do animal, e pode ser monitorado por variáveis fisiológicas e ambientais, entre elas a temperatura de superfície corporal, através da termografia por infravermelho. O objetivo deste trabalho foi desenvolver e avaliar modelos computacionais para classificação de estresse térmico em bezerras de leite por meio da termografia por infravermelho. A base de dados utilizada para a construção dos modelos foi obtida por meio de um experimento com 10 bezerras desaleitadas da raça Holandesa, alojadas em câmara climática com controle de temperatura e submetidas a condições térmicas de conforto e estresse por meio de ondas de calor, no qual foram coletados dados fisiológicos e ambientais, incluindo dados termográficos de diferentes partes do corpo do animal. Foram construídos e avaliados modelos computacionais baseados em algoritmos de aprendizado de máquina para classificação em níveis de conforto a partir de dados ambientais e de características extraídas dos dados de termografia através de um método denominado Assinatura Térmica. Para modelagem foram utilizadas ferramentas de tratamento e mineração de dados da plataforma de serviço de computação cognitiva Watson da empresa IBM. Os modelos classificadores de estresse térmico foram desenvolvidos a partir de três diferentes algoritmos de aprendizado de máquina, quatro subconjuntos da base de dados e três tipos de rotulagem de dados, e foram comparados com o método tradicional de classificação baseado no Índice de Temperatura e Umidade. Os modelos foram capazes de classificar o estresse térmico, tendo o melhor modelo obtido acurácia de 86,8%, comparável ao desempenho obtido por modelos de outros trabalhos, além de precisão de 89% para a classe de Estresse. Os resultados mostram que a abordagem baseada em aprendizado de máquina

com a assinatura térmica foi efetiva para geração de modelos com bom desempenho e com potencial de aprimoramento.

**Palavras-chave:** Bem-estar Animal; Estresse Térmico; Mineração de Dados; Modelagem Computacional; Termografia por Infravermelho.



## ABSTRACT

PEREIRA, A. L. V. **Machine learning-based model to classify calf thermal comfort using infrared thermography.** 2022. 74 f. Master thesis – Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2022.

Recent research carried out in the field of animal production relates the heat stress of animals to negative effects on their health and productivity. Heat stress is a condition that impairs the productive and reproductive performance of the animal, and can be monitored by physiological and environmental variables, including body surface temperature, through infrared thermography. The objective of this work was to develop and evaluate computational models for classifying heat stress in dairy calves using infrared thermography. The database used for the construction of the models was obtained through an experiment with 10 weaned Holstein calves, housed in a climatic chamber with temperature control and subjected to thermal conditions of comfort and stress through heat waves, in which physiological and environmental data were collected, including thermographic data from different parts of the animal's body. Computational models based on machine learning algorithms were built and evaluated to classify comfort levels from environmental data and features extracted from thermography data through a method called Thermal Signature. For modeling, data processing and mining tools from the Watson cognitive computing service platform from IBM were used. The heat stress classifier models were developed from three different machine learning algorithms, four subsets of the database and three types of data labeling, and were compared with the traditional classification method based on the Temperature and Humidity Index. The models were able to classify heat stress, and the best model obtained an accuracy of 86.8%, comparable to the performance obtained by models from other works, in addition to an accuracy of 89% for the Stress class. The results show that the approach based on machine learning with the thermal signature was effective for generating models with good performance and with potential for improvement.

**Key-words:** Animal Welfare; Heat Stress; Data Mining; Computational Modeling; Infrared Thermography.

## SUMÁRIO

1 INTRODUÇÃO .....	9
2 OBJETIVO.....	12
2.1 OBJETIVOS ESPECÍFICOS.....	12
3 HIPÓTESE .....	13
4 REVISÃO BIBLIOGRÁFICA.....	14
4.1 INFLUÊNCIA DO ESTRESSE TÉRMICO NA PRODUÇÃO ANIMAL.....	14
4.2 TECNOLOGIAS DA ZOOTECNIA DE PRECISÃO .....	17
4.3 MODELAGEM COMPUTACIONAL E MINERAÇÃO DE DADOS.....	21
<b>4.3.1 Descoberta do Conhecimento em Bases de Dados .....</b>	<b>22</b>
<b>4.3.2 Mineração de dados e aprendizado de máquina.....</b>	<b>24</b>
<b>4.3.3 Mineração de dados na agropecuária .....</b>	<b>28</b>
5 MATERIAL E MÉTODOS.....	31
5.1 DELINEAMENTO EXPERIMENTAL E COLETA DE DADOS.....	31
5.2 PRÉ-PROCESSAMENTO E SELEÇÃO DOS DADOS .....	35
5.3 EXTRAÇÃO DE CARACTERÍSTICAS.....	37
5.4 CONSTRUÇÃO DOS MODELOS.....	40
5.5 AVALIAÇÃO DOS MODELOS E FORMA DE ANÁLISE DOS RESULTADOS.....	47
6 RESULTADOS E DISCUSSÃO.....	53
6.1 DETECÇÃO DE <i>OUTLIERS</i> .....	53
6.2 ESTATÍSTICA DESCRITIVA, ANÁLISE DE CORRELAÇÃO E SELEÇÃO DE DADOS .....	54
6.3 MODELAGEM E COMPARAÇÃO DOS MODELOS .....	59
6.4 COMPARAÇÃO DOS MODELOS COM O MÉTODO TRADICIONAL....	61
6.5 DISCUSSÃO .....	64
7 CONCLUSÃO.....	67
REFERÊNCIAS.....	68

## 1 INTRODUÇÃO

Pesquisas vêm sendo recentemente realizadas no âmbito da produção animal relacionadas aos efeitos negativos do estresse térmico sobre a saúde e produtividade dos animais. Em condições de temperatura e umidade relativa elevadas o organismo do animais responde de modo a preservar a homeotermia, acionando mecanismos termorregulatórios para perda de calor, enquanto simultaneamente reduz as atividades de produção de calor. Tais mecanismos resultam em um maior gasto energético para a manutenção do equilíbrio térmico que prevalece sobre funções e comportamentos que promovem o crescimento, desenvolvimento e reprodução, impactando negativamente no desempenho produtivo e reprodutivo do animal.

Estresse térmico é causado por condições climáticas desfavoráveis ao organismo animal e sua avaliação pode ser feita pelo monitoramento de variáveis ambientais, como, por exemplo, temperatura, umidade e radiação solar, assim como pelo monitoramento de variáveis fisiológicas dos animais, como, por exemplo, frequência respiratória e temperatura interna. Alguns índices baseados em variáveis ambientais são utilizados para avaliar estresse térmico como, por exemplo, o Índice de Temperatura e Umidade (ITU) e o Índice de Temperatura de Globo Negro e Umidade (ITGU). No entanto, estes índices não medem corretamente o estresse térmico por basearem-se apenas em variáveis ambientais e não levarem em conta a resposta individual de cada animal ao seu ambiente climático.

Bons indicadores para a saúde animal e eficiência produtiva deveriam levar em conta levar em conta respostas fisiológicas como, por exemplo, a temperatura corporal, a frequência respiratória e o ganho de massa, que são próprias de cada animal. Entretanto, a medida dessas variáveis inclui tradicionalmente a coleta manual que é laboriosa e, muitas vezes, invasiva e estressante para o animal. É necessário, portanto, o desenvolvimento de sistemas para avaliação do estresse térmico a partir de fatores ambientais e fisiológicos que sejam mais práticos e menos invasivos. Tais sistemas podem viabilizar o manejo adequado do rebanho, contribuindo significativamente para tomada de decisões que resultem no aumento de produtividade e no uso eficiente de recursos envolvidos no sistema de produção.

Uma solução que vem sendo estudada para este problema é a utilização de dados obtidos por Termografia de Infravermelho (TIV) (MCMANUS, C. et al., 2016). Esta técnica consiste na captação da radiação eletromagnética emitida pela superfície corporal do animal, na faixa do infravermelho, e sua conversão em uma imagem térmica, em que a cada pixel está associado um valor de temperatura. Por se tratar de uma medição sem contato físico e passível de automação, a TIV pode ser usada como método não invasivo para avaliação de estresse térmico. Salienta-se que se chama frequentemente de imagem termográfica ou térmica, mas especificamente, trata-se de uma matriz de dados de temperatura que pode ser representada visualmente através de uma escala de cores associadas aos valores de temperatura. Neste trabalho serão utilizados como sinônimos as expressões imagem termográfica e matriz de temperaturas para se referir a dados de TIV, de forma que cada expressão será utilizada de acordo com o contexto para melhor clareza dos aspectos abordados em cada tópico.

A medição indireta por imagem termográfica geralmente demanda a extração de características dessa imagem e a aplicação destas a um modelo computacional que seja capaz de prever a resposta do animal às condições ambientais. No caso de sistemas biológicos, a grande complexidade das interações entre variáveis dificulta os processos de modelagem e predição. Neste contexto, técnicas baseadas em aprendizado de máquina (*machine learning*) têm alcançado resultados promissores para geração de modelos preditivos para avaliação de estresse térmico utilizando sensores não invasivos (MCMANUS, C. et al., 2016).

Diversas plataformas e softwares, pagos ou gratuitos, oferecem ambientes virtuais onde é possível desenvolver modelos computacionais para predição ou classificação. O Watson Studio é uma plataforma de serviços de computação cognitiva desenvolvida pela IBM (International Business Machines Corporation, EUA) e que oferece várias ferramentas e algoritmos de aprendizado de máquina e mineração de dados. Esta plataforma apresenta-se como uma possível alternativa para o problema em questão por oferecer muitas opções de tratamento e modelagem de dados, além de possuir uma versão gratuita.

Diante deste cenário, identificam-se oportunidades de pesquisa visando o desenvolvimento de classificadores automáticos de estresse térmico para a

produção animal através de medições não invasivas aplicadas a modelos de predição ou classificação baseados em aprendizado de máquina.

## 2 OBJETIVO

Desenvolver e avaliar modelos computacionais baseados em técnicas de mineração de dados para classificação de estresse térmico em bovinos de leite, utilizando dados ambientais e a temperatura da superfície corporal obtida por meio de termografia de infravermelho.

### 2.1 OBJETIVOS ESPECÍFICOS

- Construir base de dados a partir de experimento com bezerras de leite conduzido em câmara climática;
- Utilizar a base de dados para desenvolver método de extração de características e para modelagem computacional;
- Explorar a plataforma de serviços de computação cognitiva Watson Studio como um ferramenta para modelagem das respostas biológicas relacionadas com o estresse térmico;
- Avaliar ferramentas de aprendizado de máquina e evidenciar as que obtiverem melhor potencial para construção de modelos computacionais de classificação da condição térmica de conforto e estresse;
- Elaborar procedimento de construção de modelos computacionais de classificação de estresse térmico.

### **3 HIPÓTESE**

Espera-se que modelos computacionais desenvolvidos a partir de técnicas de mineração de dados sejam capazes de classificar o estresse térmico de bezerras de leite a partir de dados não-invasivos com desempenho superior aos métodos clássicos de avaliação de conforto térmico.

## 4 REVISÃO BIBLIOGRÁFICA

A pecuária leiteira brasileira se destaca no cenário internacional. O Brasil foi, segundo apontam dados da FAO (2022), o 3º maior produtor de leite do mundo em 2020, com 36,5 bilhões de litros produzidos, aproximadamente 5% da produção mundial, atrás apenas de EUA e Índia, com 101,3 e 87,8 bilhões de litros respectivamente. Dados do Instituto Brasileiro de Geografia e Estatística (2022) apontam que em 2021 a produção brasileira de leite foi de 25,1 bilhões.

No período entre 2008 e 2016, houve um expressivo crescimento da produção de leite no Brasil. O volume de leite produzido aumentou 21,89% no período, enquanto que a média mundial foi de 13,27%. Em termos de produtividade, estimada em litros por vaca por ano, o Brasil teve o maior ganho de produtividade no mundo, de 33,7%, sendo a média mundial de 4,5% segundo dados da Companhia Nacional de Abastecimento – Conab (2018).

Apesar dos avanços recentes, o Brasil ainda possui produtividade de leite baixa se comparada com outros países, estando na 14ª posição, com produtividade abaixo da média mundial (CONAB, 2018). A projeção do Ministério da Agricultura (BRASIL, 2020) estima um aumento de 21,3% na produção no período de 2019/20 até 2029/30, que deve se dar mais devido a melhorias na gestão das fazendas e na produtividade dos animais e menos devido ao número de vacas em lactação.

Diante do cenário de necessidade de melhorias na gestão e aumento de produtividade da pecuária leiteira, tem-se realizado diversos estudos acerca dos efeitos do estresse térmico nos animais. O estresse térmico provocado por alta temperatura e umidade relativa do ar impacta negativamente na saúde e em aspectos produtivos e reprodutivos dos animais, reduzindo o volume produzido e a qualidade do leite. No Brasil, onde o clima é predominantemente tropical, as condições meteorológicas muitas vezes não proporcionam um ambiente de conforto térmico para os animais, o que torna necessária a realização de estudos para aprofundar o conhecimento sobre o estresse térmico e meios para monitorá-lo e controlá-lo (CARVAJAL et al., 2021).

### 4.1 INFLUÊNCIA DO ESTRESSE TÉRMICO NA PRODUÇÃO ANIMAL



Estresse térmico pode ser definido como a soma das forças internas e externas que atuam no animal e causam um aumento na temperatura corporal e provocam uma resposta fisiológica (BECKER et al., 2020). Em animais homeotérmicos, como é o caso dos bovinos, há a necessidade de manter a temperatura corporal em um valor relativamente constante para o correto funcionamento do organismo. O aumento da temperatura corporal além deste valor ideal, que pode ser causado por fatores externos tais como temperatura do ar, umidade relativa e radiação solar, provoca uma resposta termorregulatória do organismo no sentido de dissipar calor, visando manter a homeotermia.

Os mecanismos de termorregulação classificam-se em dois tipos, os de via evaporativa (transferência de calor latente) e de via não evaporativa (condução, convecção e irradiação), e podem ocorrer como comportamentos ou como reações fisiológicas. Condições de temperatura do ar mais elevadas desencadeiam respostas de via evaporativa, enquanto que temperaturas mais amenas (condições de estresse térmico não muito intensas) desencadeiam respostas de via não evaporativa. Mecanismos de via não evaporativa incluem contato físico com superfícies que estejam a temperaturas mais baixas que a temperatura corporal, exposição a ambientes com temperatura do ar mais baixa que a temperatura corporal e menor exposição à radiação solar. Estes mecanismos não requerem gasto de energia. Dissipação de calor por via evaporativa, que inclui ofego e sudorese, por outro lado, envolve gasto de energia, e sua eficiência é maior quanto menor for a umidade relativa do ar (FERREIRA et al., 2017).

Ocorre que a termorregulação dos animais homeotérmicos é uma reação que impacta negativamente no desempenho produtivo e reprodutivo, devido ao fato de exigir redução do metabolismo e gasto de energia. Atividades como produção de leite, ingestão de alimentos e crescimento são reduzidas a fim de diminuir a geração de calor metabólico pelo organismo, enquanto que mecanismos termorregulatórios que envolvem gasto de energia são acionados, resultando em menos energia direcionada para atividades produtivas. Efeitos do estresse térmico em vacas leiteiras são bem documentados, podendo provocar piora na performance reprodutiva, perda de apetite e conseqüente redução na ingestão de alimento, redução na produção de leite, alterações no teor de

gordura e proteína do leite e depressão das funções imunológicas, o que facilita o desenvolvimento de doenças (BECKER et al., 2020).

Vacas leiteiras apresentam máximo potencial genético com mínimo custo fisiológico e máxima produtividade apenas dentro da zona termoneutra. Zona termoneutra é definida como a faixa de temperaturas do ambiente na qual o animal não precisa gastar energia para manter a temperatura corporal normal. Para vacas em lactação, a zona termoneutra está entre 5 a 25 °C, podendo variar em função da idade, raça, produção, condição corporal, estágio da lactação, prenhez, tipo de cobertura capilar e pigmentação da pele e do pelo. Alguns autores consideram 24 °C o valor crítico para estresse térmico em vacas lactantes (JOHNSON; VANJONACK, 1976; MARTELLO, 2006; NÃÃS, 1989).

Além de aspectos produtivos e reprodutivos, o bem-estar animal também é negativamente afetado pelo estresse térmico. Polsky e Von Keyserlingk (2017) analisaram o efeito do estresse térmico em vacas leiteiras com base no modelo de três chaves construtivas do bem estar animal: saúde e função biológica, estados afetivos e naturalidade de seu comportamento. Os autores afirmam que o estresse térmico afeta negativamente os estados afetivos do animal, provocando sentimentos de fome e sede, e potencialmente frustração, agressividade e dor.

Tradicionalmente, tem-se usado índices climáticos como forma de avaliação do conforto térmico dos animais. O Índice de Temperatura e Umidade (ITU) e o Índice de Temperatura de Globo Negro e Umidade (ITGU) estão entre os mais utilizados. O primeiro é uma combinação da temperatura e umidade relativa do ar e o segundo considera, além da temperatura e umidade, a radiação solar e o ponto de orvalho. Estes índices são úteis para avaliar o conforto térmico por se basearem em informações de fácil acesso por instrumentos de medida ou estações meteorológicas, sem a necessidade de contato com os animais. Entretanto, essas medidas consideram apenas variáveis ambientais, sem considerar variáveis fisiológicas dos animais, o que dificulta a avaliação do estresse térmico a nível individual. Bem estar é uma característica relativa a animais enquanto indivíduos, a preferência deve ser dada a medidas válidas de bem estar individual (WINCKLER, 2019).

A resposta individual de um animal ao estresse térmico pode ser melhor aferida através de medidas fisiológicas, estando a frequência respiratória (FR) e

a temperatura retal (TR) entre as mais usadas para avaliar a carga de calor em vacas leiteiras (HOFFMAN et al., 2019). Os valores de temperatura retal entre 38,0 a 39,3 °C e frequência respiratória abaixo de 60 mov/min são consideradas adequados para bovinos de leite adultos (DUKES, 1996; HAHN; MADER, 1997). No entanto a aferição da FR e da TR mostra-se um processo lento, invasivo e laborioso no ambiente produtivo, sendo impraticáveis para produções em larga escala.

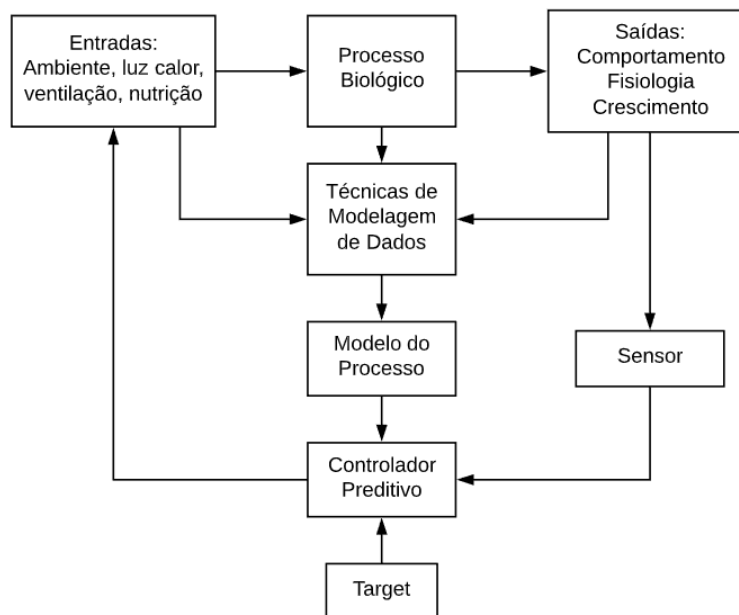
Encontrar maneiras de monitorar o estresse térmico de animais de produção que sejam ao mesmo tempo práticas, não-invasivas e individualizadas mostra-se um desafio. Tecnologias modernas tais como sensores, modelos computacionais e sistemas automatizados tem sido pesquisadas para buscar uma solução. Nesse contexto, a Zootecnia de Precisão emerge como um paradigma de uso da tecnologia na gestão da produção animal.

## 4.2 TECNOLOGIAS DA ZOOTECNIA DE PRECISÃO

Zootecnia de Precisão pode ser definida como o gerenciamento da produção animal utilizando os princípios e tecnologias da engenharia de processos (WATHES et al., 2008). A zootecnia de precisão enxerga a pecuária como um conjunto de processos interligados que podem ser compreendidos como um sistema.

O fluxograma da Figura 1 mostra o funcionamento geral de um sistema de produção baseado na zootecnia de precisão. O sistema é constituído pelas seguintes etapas: (1) o processo biológico (o animal), que é afetado pelo ambiente (*inputs*), que são variáveis externas tais como temperatura, umidade e nutrição; (2) sensores, que medem as variáveis ambientais e os sinais de saída (*outputs*) do processo, tais como comportamento e variáveis fisiológicas; (3) um modelo computacional capaz de prever a evolução dinâmica do processo, que foi previamente gerado a partir de uma base de dados e técnicas de modelagem; (4) o valor desejado para a saída do processo (*target*), que representa os objetivos da produção, e (5) um controlador que utilize o modelo computacional e os dados dos sensores para acionar os atuadores que controlam os *inputs* do processo, de acordo com o *target* definido (WATHES et al., 2008).

Figura 1 – Fluxograma de sistema de produção baseado na zootecnia de precisão



Fonte: WATHES, C. M. et al. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? **Computers and Electronics in Agriculture**, Smart Sensors in precision livestock farming. v. 64, n. 1, p. 2–10, 1 nov. 2008.

Wathes et al. (2008) afirmam que a Zootecnia de Precisão permite que o manejo dos animais de produção seja feito a nível individual, mas apenas se for possível mensurar os sinais de saída de cada animal individualmente e se isso se justificar em termos econômicos. O manejo individualizado dos animais é desejável, visto que cada animal possui suas particularidades e responde de maneira diferente às variáveis ambientais. Ao ajustar o manejo de acordo com as necessidades particulares de cada animal, pode-se aproveitar ao máximo o potencial genético de cada um. Entretanto, dado que é necessário mensurar os sinais de saída de cada animal, é necessária a pesquisa e desenvolvimento de sensores e técnicas de sensoriamento que permitam atingir este objetivo de maneira eficiente.

Na criação de vacas leiteiras, diferentes abordagens têm sido estudadas para aferir a nível individual o estado no qual se encontra o animal. No monitoramento da saúde animal, pode-se citar o uso de análise de imagens para detecção de lesões no joelho de vacas (GUO; HE; SONG, 2018) e uso de análise de vídeo para detecção de claudicação (ABDUL JABBAR et al., 2017; VAN HERTEM et al., 2014). No monitoramento do comportamento, pode-se citar o

uso de acelerômetros e outros sensores para detecção de estro e de trabalho de parto (ARCIDIACONO; MANCINO; PORTO, 2020; BENAÏSSA et al., 2020; KRIEGER et al., 2018), o uso de diversos sensores para estimação de ingestão de água (WILLIAMS et al., 2020), e a estimação de ingestão de alimento através do monitoramento do movimento da mandíbula, feito por meio de sensores de pressão (PAHL et al., 2016), sensores acústicos (NAVON et al., 2013) e acelerômetros (MATTACHINI et al., 2016; OUDSHOORN et al., 2013). Também pode-se citar o uso de visão computacional para estimação de altura e peso corporal de bezerras de leite (NIR et al., 2018).

Entre as diversas técnicas de sensoriamento existentes, destaca-se a termografia de infravermelho como uma técnica capaz de medir a temperatura de uma superfície à distância. Esta técnica baseia-se no fato de que qualquer objeto que esteja a uma temperatura acima do zero absoluto emite calor na forma de radiação na região do infravermelho do espectro eletromagnético. Com o uso de câmeras termográficas é possível detectar este tipo de radiação mesmo em mínimas variações (MCMANUS et al., 2016).

No âmbito da produção animal, a termografia de infravermelho é definida como um método de sensoriamento remoto não-invasivo usado na mensuração de mudanças na transferência de calor e no fluxo sanguíneo através da detecção de pequenas mudanças na temperatura corporal (MCMANUS et al., 2016). A medição da temperatura dos animais por meio de câmeras termográficas traz uma grande vantagem se comparadas à medição por meio de termômetros convencionais por possuírem facilidade de utilização, não exigirem contenção e contato físico com os animais e serem passíveis de automatização. Knizkova et al. (2007) afirmam que a termografia de infravermelho é uma técnica não invasiva, capaz de captar o calor emitido através da superfície dos animais.

O monitoramento da temperatura de superfície corporal dos animais permite avaliar alguma situação fisiológica anormal no animal. Stelletta et al. (2012) afirmam que como a temperatura de superfície reflete o estado do metabolismo do tecido do animal e da circulação sanguínea, alterações nos padrões de temperatura podem significar a presença de inflamações ou alterações circulatórias.

Estudos têm sido feitos visando identificar possíveis aplicações da termografia na detecção de processos inflamatórios e diagnóstico de doenças.

Berry et al. (2003) observaram a temperatura do úbere de vacas leiteiras e concluíram que a termografia é um método promissor para detecção de mastite; enquanto que Polat et al. (2010), concluíram que a termografia pode ser empregada para detecção de mastite subclínica, mas sua confiabilidade para vacas com diferentes características e em diferentes ambientes ainda precisa ser determinada. Martins et al. (2013) mostraram que imagens termográficas podem ser usadas para detecção de mastite em ovelhas, inclusive em estágio subclínico. Outros estudos incluem detecção de infecção sistêmica (SCHAEFER et al., 2004), febre aftosa (RAINWATER-LOVETT et al., 2009), laminitite (NIKKHAH et al., 2005) e dermatite digital (ALSAAOD et al., 2014).

Outra possível aplicação da termografia por infravermelho seria na avaliação do estado emocional de animais. Uddin et al. (2021) buscaram uma possível associação entre temperatura da superfície corporal e a resposta comportamental à ansiedade em vacas leiteiras. Os autores coletaram dados de temperatura retal, imagens termográficas do olho, banda coronária e membros anteriores dos animais e realizaram testes indicadores de ansiedade. Nesse estudo concluíram que mensuração por termografia de infravermelho pode ser um indicador de emoções nas vacas, sendo necessários estudos mais detalhados.

O uso de imagens termográficas também têm sido estudado para a avaliação de estresse térmico. Daltro et al. (2017) investigaram se a termografia de infravermelho é útil para o reconhecimento de vacas leiteiras em estado de estresse térmico, bem como identificar a melhor região do corpo do animal para fazer essa avaliação, e concluíram que a termografia é eficaz para a determinação do estresse térmico, sendo a região lateral do úbere a que apresentou a maior correlação (0,74) com a temperatura retal. Stewart et al. (2017) validaram um método de mensuração indireta da frequência respiratória, no qual grava-se um vídeo em termografia das narinas das vacas e em seguida faz-se a contagem manual do número de respirações, observadas pela menor temperatura das narinas durante a inspiração e maior temperatura durante a expiração.

Entretanto, existem algumas limitações ao uso da termografia por infravermelho no monitoramento de temperatura de animais. Segundo McManus et al. (2016), a incidência de radiação solar direta, a alta umidade, a perda de

calor devido ao vento e eventual sujeira sobre a superfície imageada afetam as medidas de temperatura. Também deve-se ressaltar que a radiação captada pelas câmeras não depende apenas da temperatura da superfície, mas também de sua emissividade e condutividade térmicas. Church et al. (2014) também apontam para a necessidade de considerar potenciais erros na mensuração da temperatura a partir de câmeras termográficas, que podem ser causados por diferenças na distância da câmera ao ponto focal, velocidade do vento e carga de radiação solar, além de ser preciso desenvolver uma nova metodologia para medir com maior precisão o valor da emissividade da pele e do pelo dos animais.

Além de limitações particulares da termografia, há de se considerar também desafios em um contexto mais geral da zootecnia de precisão. À medida que são desenvolvidas e implementadas diferentes técnicas de sensoriamento para monitoramento de animais, cresce o volume e a complexidade dos dados gerados pelos sensores. Astill et al. (2020) discutiram o impacto da adoção de sensores na avicultura e apontaram para a necessidade de se estabelecer sistemas de coleta, armazenamento e acesso de dados que maximizem a utilidade dessas tecnologias, bem como a importância de técnicas de análise de grandes volumes de dados (*Big Data*) para melhor aproveitar o potencial das tecnologias da Zootecnia de Precisão. Diante destes desafios, é necessário adotar uma metodologia para extrair informações a partir de grandes volumes de dados e, nesse sentido, o processo de Descoberta do Conhecimento em Bases de Dados (*KDD – Knowledge Discovery in Databases*) mostra-se como uma ferramenta promissora para o enfrentamento deste problema.

#### 4.3 MODELAGEM COMPUTACIONAL E MINERAÇÃO DE DADOS

Tradicionalmente, a análise de dados visando a identificação de regras ou padrões é feita de forma manual. Diante de um fenômeno de interesse, faz-se uma coleta de dados em condições controladas de tal maneira que fiquem evidentes as informações necessárias para a sua compreensão. Um especialista então analisa os dados buscando reconhecer eventuais padrões que possam ser usados para melhor descrever os mecanismos que atuam no fenômeno, prever variáveis desconhecidas ou o seu comportamento futuro e prescrever alguma

recomendação para que se interfira no sistema visando alguma resposta desejada.

Entretanto, com o avanço da tecnologia, cada vez mais encontram-se fenômenos que são muito complexos ou que contenham um grande volume de dados, sendo a análise manual destes excessivamente lenta, cara, e subjetiva em diversas aplicações. Em vários domínios, a análise manual não é mais prática devido ao crescimento dramático observado no volume de dados (FAYYAD; PIATETSKY-SHAPIRO; SMYTH, 1996).

As bases de dados podem ter crescimento de duas formas: o crescimento do número de objetos ou observações (registros), e o crescimento do número de variáveis (atributos) de cada objeto. Fayyad, Piatetsky-Shapiro e Smyth (1996), pontuaram que cada vez mais são frequentes bases de dados com um número de registros da ordem de centenas de milhões, como nas ciências astronômicas, e número de atributos da ordem de dezenas ou até mesmo de centenas, como no diagnóstico médico. É impraticável analisar manualmente dados dessa magnitude, algum tipo de automação mostra-se necessária.

A solução que vem sendo adotada nas diferentes áreas que lidam com este problema tem sido o uso de computadores. Graças aos avanços da computação, pode-se processar grandes volumes de dados em alta velocidade de maneira automática. Com o problema da sobrecarga de dados e as novas ferramentas disponíveis para resolvê-lo, há a necessidade de desenvolver uma metodologia para utilizar tais ferramentas de maneira sistematizada. Nesse sentido, destaca-se o KDD entre os métodos de referência para o processo de descoberta do conhecimento a partir base de dados brutos.

#### **4.3.1 Descoberta do Conhecimento em Bases de Dados**

O KDD é um método que foi proposto e detalhado por Fayyad, Piatetsky-Shapiro e Smyth (1996) e enfatiza o conhecimento como produto final do processo de descoberta. Entretanto o KDD não é um método definitivo, mas um dos métodos de referência para orientar ou sistematizar o processo de descoberta de conhecimento, que pode ser adequado ou associado a outros métodos de acordo com as demandas impostas pelas características da base de dados e das informações preteridas. Diversas áreas de conhecimento compõe o



KDD, tais como aprendizado de máquina, reconhecimento de padrões, bancos de dados, estatística, inteligência artificial, aquisição de conhecimento para sistemas especialistas, visualização de dados e computação de alta performance.

Fayyad, Piatetsky-Shapiro e Smyth (1996) definem KDD como “o processo não-trivial de identificação de padrões válidos, novos, potencialmente úteis e em última instância compreensíveis em dados”. Por não-trivial, entende-se que o processo não consiste de uma sequência de comandos fixos, mas envolve algum tipo de busca ou inferência. Os padrões identificados no KDD são algum tipo de descrição de alto nível do conjunto de dados. Tais padrões são considerados de fato conhecimento se atingirem critérios pré-estabelecidos de validade, novidade, utilidade e compreensibilidade.

O KDD é um processo constituído de 5 etapas:

1. Seleção, na qual são selecionados quais dados irão compor a base de dados a ser analisada;
2. Pré-processamento, no qual faz-se a limpeza dos dados, remoção de ruído, elaboração de estratégias para lidar com dados que tenham atributos faltantes, entre outros;
3. Transformação, também chamada de extração de características, na qual faz-se a redução de dimensionalidade através da procura de características que representem o sistema melhor do que os dados brutos;
4. Mineração de dados, na qual determina-se o método de mineração de dados que melhor atende aos objetivos previamente definidos e escolhe-se o algoritmo e os parâmetros mais apropriados, obtendo assim modelos e/ou padrões que de alguma forma representam o conhecimento decoberto, seja por meio de regras de decisão, fórmulas ou agrupamentos;
5. Interpretação/avaliação, na qual os modelos e/ou padrões encontrados são visualizados, interpretados e têm seu desempenho avaliado por meio de alguma métrica, podendo inclusive ser necessário o retorno a alguma das etapas anteriores para aprimoramento.

Uma vez desenvolvido um modelo satisfatório, pode-se implementá-lo em alguma aplicação.

#### **4.3.2 Mineração de dados e aprendizado de máquina**

Fayyad, Piatetsky-Shapiro e Smyth (1996) definem mineração de dados como “a etapa do processo do KDD que consiste em aplicar análises de dados e algoritmos de descoberta que, dentro de limitações aceitáveis de eficiência computacional, produzem uma enumeração particular de padrões (ou modelos) dos dados”.

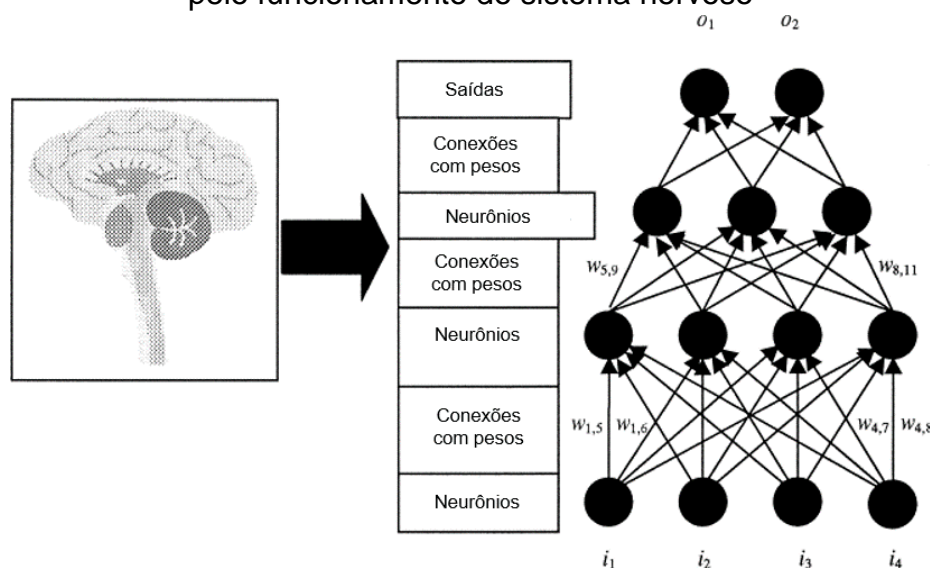
A mineração de dados pode ser feita seguindo alguns métodos distintos, de acordo com a finalidade da análise. Os algoritmos de aprendizado de máquina utilizados em mineração de dados classificam-se segundo a finalidade em algoritmos de aprendizagem supervisionada e de aprendizagem não supervisionada. A aprendizagem supervisionada é um processo no qual um algoritmo utiliza como entrada uma base de dados na qual cada registro está rotulado, ou seja, a saída desejada para eles está previamente definida. Desenvolve-se então um modelo que seja capaz de prever rótulos desconhecidos para novos registros. A aprendizagem não supervisionada consiste na identificação de padrões em bases de dados sem que haja conhecimento prévio de tais padrões. De forma geral, os métodos de aprendizagem supervisionada são métodos preditivos do comportamento do sistema, enquanto os de aprendizagem não supervisionada são métodos descritivos do sistema.

No caso da aprendizagem não supervisionada, os algoritmos dividem-se em algoritmos de associação, agrupamento, padrões sequência, detecção de anomalias e sumarização. Já na aprendizagem supervisionada os algoritmos se dividem-se em algoritmos de classificação e algoritmos de regressão (ou predição). Na classificação, é atribuída a cada registro da base de dados uma classe discreta (rótulo ou saída do modelo) e, após o treinamento, o modelo é capaz de classificar novos dados em uma das classes. Na predição, a saída do modelo (atributo meta) é um valor numérico contínuo.

Dentre os principais algoritmos de classificação, destacam-se três: redes neurais artificiais (RNA), *support vector machine* (SVM) e árvores de decisão.

Redes Neurais Artificiais são um algoritmo de aprendizado de máquina inspirado pelo funcionamento do sistema nervoso, como retratado na Figura 2, constituindo-se de uma rede de elementos de processamento, denominados neurônios computacionais ou perceptrons, que simula dentro de limites a rede de neurônios do cérebro e sua atividade bio-elétrica. Os neurônios são organizados em camadas, nas quais cada um deles é conectado com neurônios das camadas anterior e posterior, simulando as conexões sinápticas do sistema nervoso. Os valores de saída dos neurônios de uma camada são utilizados como entrada pelos neurônios da camada subsequente. É associado a cada conexão entre neurônios um valor denominado peso. A Figura 3 ilustra o modelo de perceptron e seu funcionamento em rede. Um neurônio recebe como entrada a soma dos valores de saída dos neurônios da camada anterior multiplicados pelos respectivos pesos, mais uma entrada extra denominada *bias*. É aplicada a esta soma uma função de ativação  $\varphi(v)$ , que fornece então o valor de saída do neurônio. Diversas funções são usadas como funções de ativação, sendo as mais comuns as funções sigmoide, ReLU e linear. O processamento da informação pelas camadas de neurônios se repete até a última camada, que fornece a saída final da rede, e pelo ajuste dos bias e dos pesos de cada conexão ocorre o aprendizado da rede como um todo (WALCZAK; CERPA, 2003).

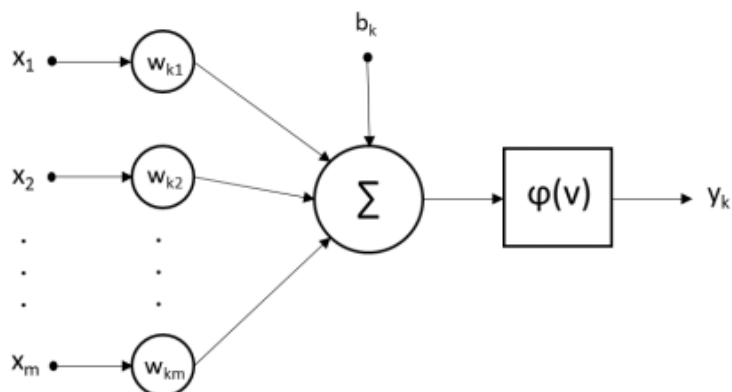
Figura 2 – Exemplo da arquitetura de uma rede neural artificial, que é inspirada pelo funcionamento do sistema nervoso



Fonte: WALCZAK, S.; CERPA, N. Artificial Neural Networks. In: **Encyclopedia of Physical Science and Technology**. [s.l.] Elsevier, 2003. p. 631–645.

Legenda:  $i$  – entradas da rede neural;  $w$  – pesos das conexões;  $o$  – saídas da rede neural.

Figura 3 – Modelo do funcionamento de um neurônio artificial



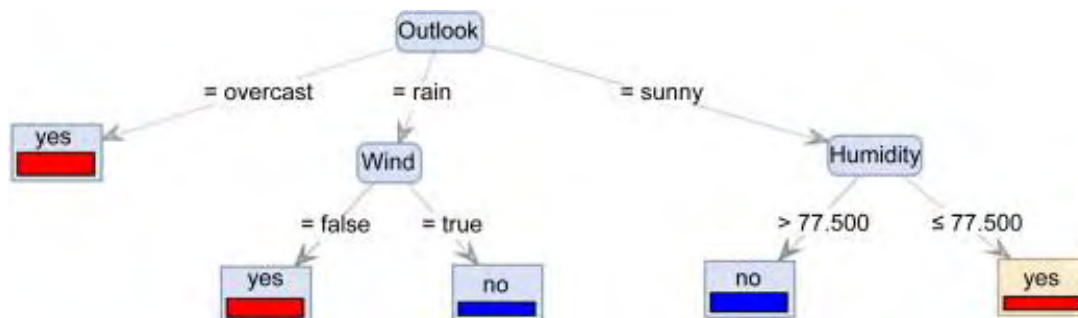
Fonte: HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2. Ed. Tradução de Paulo Martins Engel. São Paulo: Editora Artmed LTDA, 1999.

Legenda:  $x$  – entradas do neurônio artificial;  $w_k$  - pesos que multiplicam as entradas;  $b_k$  - bias;  $\varphi(v)$  – função de ativação;  $y_k$  – saída do neurônio artificial

RNA's podem ser utilizadas tanto como um método de aprendizagem supervisionada como aprendizagem não supervisionada, tendo algoritmos de treinamento adequados para cada caso.

Árvores de decisão são algoritmos utilizados para classificação ou predição que se estruturam como um fluxograma de decisões tomadas a partir dos atributos preditores a fim de classificar um dado de entrada ou predizer um valor. A decisão pode ser tomada a partir de atributos numéricos ou categóricos. Cada nó representa uma regra de decisão acerca de um atributo preditor e os ramos representam possíveis valores desse atributo. As folhas (nós terminais) representam o fim do processo de decisão, que retorna a classe a ser atribuída para o dado de entrada. A Figura 4 mostra um exemplo de árvore de decisão.

Figura 4 – Exemplo de um modelo de árvore de decisão



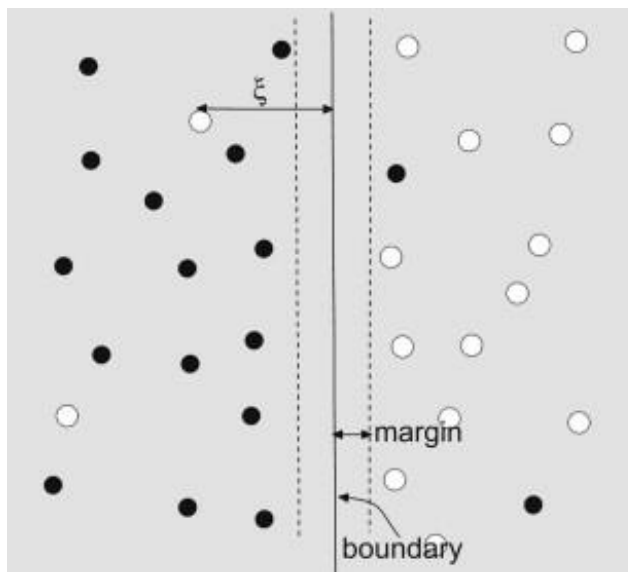
Fonte: KOTU, V.; DESHPANDE, B. Classification. In: **Predictive Analytics and Data Mining**. [s.l.] Elsevier, 2015. p. 63–163.

Na Figura 4, cada nó da árvore de decisão indica uma decisão do modelo acerca do dado de entrada, sendo a primeira decisão sobre a condição meteorológica (nublado, chuvoso ou ensolarado) e as outras decisões sobre a presença ou não de vento e a umidade. O treinamento de um modelo de árvore de decisão se baseia na maximização do ganho de informação obtido por cada uma das regras de decisão. Dada uma base de dados, o algoritmo de treinamento escolhe um atributo preditor e calcula o ganho de informação obtido ao se criar uma regra de decisão a partir deste atributo. O ganho de informação representa o quão valioso um atributo é para prever a classe de um dado de entrada, e é calculado a partir da entropia de informação associada ao atributo. O algoritmo repete o procedimento para cada atributo e a regra será criada para o atributo com maior ganho de informação. As próximas regras são criadas com o mesmo processo até que não haja mais ganho de informação significativo com a criação de novas regras ou a árvore tenha atingido um limite de tamanho previamente definido (KOTU; DESHPANDE, 2015).

O algoritmo SVM (*Support Vector Machine*, Máquina de Vetores de Suporte) é um algoritmo de classificação que se baseia na criação de uma fronteira de separação entre dois conjuntos de dados a partir de seus atributos preditores para separar dados de diferentes classes. Para definir a melhor fronteira possível, é calculada a distância que separa os pontos da fronteira, denominada margem (Figura 5). Uma vez definida a fronteira, pode-se classificar dados de entrada do modelo verificando de qual lado da fronteira os dados se encontram (KOTU; DESHPANDE, 2015).

O treinamento de um modelo de SVM é feito através da minimização do erro (representado na Figura 5 pelo símbolo  $\xi$ ), definido como a distância dos pontos erroneamente classificados até a margem (*margin*) em relação a fronteira entre os conjuntos de dados (*boundary*). Para casos de dados que não sejam linearmente separáveis, é possível aplicar funções de transformação sobre o espaço de pontos, denominadas kernel, a fim de reorganizar os dados, tornando possível separá-los por uma fronteira linear.

Figura 5 – Ilustração gráfica do funcionamento do algoritmo SVM



Fonte: KOTU, V.; DESHPANDE, B. Classification. In: **Predictive Analytics and Data Mining**. [s.l.] Elsevier, 2015. p. 63–163.

Legenda:  $\xi$  – erro da estimativa da margem; *Boundary* – fronteira de separação dos dados; *Margin* – margem adotada para separação dos dados.

#### 4.3.3 Mineração de dados na agropecuária

Com o avanço da tecnologia de sensoriamento discutida na Seção 2.2, e as soluções computacionais para os problemas decorrentes desse avanço, tem surgido nos últimos anos um paradigma organização da produção agropecuária denominado *Smart Farming*.

*Smart Farming* é um paradigma que surgiu a partir da incorporação de diversas tecnologias de informação e comunicação na agropecuária, tais como sensoriamento remoto, computação em nuvem e Internet das Coisas, com a finalidade de aprimorar o processo de manejo e tomada de decisão (KAMILARIS; KARTAKOULLIS; PRENAFETA-BOLDÚ, 2017).

Apesar dos grandes avanços obtidos no âmbito da agricultura, na pecuária a produção acadêmica não é tão rica. Devido à própria natureza dos animais, que, diferentemente dos vegetais, são seres que se movimentam e têm maior variabilidade de características físicas e comportamentais, existem dificuldades maiores no desenvolvimento de soluções envolvendo sensoriamento remoto e mineração de dados. Morota et al. (2018) publicaram uma revisão sobre aprendizado de máquina e mineração de dados na produção

animal, na qual mencionam aplicações em estudos de genética, detecção de dados fora do padrão (*outliers*) em bases de dados, imputação genotípica, detecção de mastite, estimação de peso corporal e microbiota ruminal. Neethirajan (2020) estudou como os conhecimentos em sensores, *Big Data*, inteligência artificial e aprendizado de máquina contribuem para a redução de custos de produção, aumento de eficiência, melhoria do bem-estar animal e aumento da produção na pecuária.

Os avanços em técnicas de mineração de dados e aprendizado de máquina possibilitam que grandes bases de dados sejam analisados com maior rapidez. Entre as áreas que se beneficiam deste progresso, está o monitoramento do estresse térmico de animais. Os efeitos prejudiciais do estresse térmico em animais já foram discutidos na Seção 2.1.

Hoffmann et al. (2019) publicaram uma revisão onde analisam técnicas não-invasivas para determinação de estresse térmico em vacas leiteiras, que são baseados em indicadores fisiológicos, comportamentais ou produtivos. Os autores citam a termografia do infravermelho como um bom método para avaliação do conforto térmico. Ricci, Silva-Miranda e Titto (2019) estudaram a termografia do infravermelho como um método não-invasivo para avaliação de estresse térmico em suínos e concluíram que ela pode ser uma ferramenta para avaliar o ambiente térmico de instalações para suínos e o bem-estar animal. Giro et al. (2019) estudaram o uso de termografia para monitoramento de temperatura corporal de gado de corte a pasto, tendo concluído que é uma tecnologia segura e não-invasiva para inferir a temperatura interna. Salles et al. (2016) estudaram como diversas regiões da superfície corporal podem ser analisadas com termografia em vacas leiteiras da raça Jersey em situação de conforto térmico, tendo encontrado que a região da frente possui maior associação com a temperatura retal. Dias Batista et al. (2021) desenvolveram modelos de classificação de estresse térmico de bezerros girolando da região do semiárido brasileiro a partir de imagens termográficas, utilizando análise geoestatística e multivariada, tendo os modelos conseguido ser eficientes na caracterização do estresse térmico com um pequeno número de atributos preditores. Becker et al. (2021) utilizou 3 algoritmos (regressão logística, floresta randômica e Naïve Bayes gaussiano) com diversas variáveis fisiológicas e comportamentais e 3 tratamentos distintos (estrutura de sombra portátil, sistema portátil de aspersão

e controle sem abatimento de calor), obtendo melhores resultados para o algoritmo de floresta randômica.

Um dos pontos positivos da termografia do infravermelho, se comparada com a análise manual, é a possibilidade de automatização do processo de monitoramento da condição de conforto térmico dos animais. Tal automatização pode ser realizada através de modelos preditivos baseados em algoritmos de aprendizagem supervisionada. Sousa et al. (2016) desenvolveram um classificador baseado em lógica fuzzy para avaliar estresse térmico de gado de corte em confinamento a partir de imagens termográficas e dados meteorológicos, que obteve uma performance superior à avaliação através do índice de temperatura e umidade (ITU). Os autores concluíram que o uso de termografia associado a sistemas de inferência fuzzy mostram grande potencial para predição de temperatura retal. Sousa et al. (2018), na mesma linha que o trabalho anterior, utilizaram RNA's para análise das imagens termográficas, tendo o modelo proposto obtido boa performance, se comparado com o modelo clássico de regressão e uma maior capacidade de generalização. Pacheco et al. (2020) estudaram o uso de RNA's na construção de um classificador de estresse térmico para vacas leiteiras da raça Holandesa, também obtendo melhores resultados que os métodos tradicionais de índice de temperatura e umidade (ITU) e de índice de temperatura de globo negro e umidade (ITGU).

Diante dos trabalhos recentemente publicados, verifica-se o potencial do uso de algoritmos de inteligência artificial, em especial de aprendizado de máquina como uma ferramenta de análise para avaliar o nível de estresse térmico dos animais a partir de imagens termográficas, o que demonstra a necessidade de mais estudos acerca destas ferramentas. Diferentes algoritmos podem ser testados para avaliar o estresse térmico, que podem obter resultados melhores que os anteriormente estudados.



## 5 MATERIAL E MÉTODOS

Os modelos classificadores de estresse térmico foram criados a partir de uma base de dados construída com um experimento de campo. A seguir está descrito o delineamento experimental, assim como as etapas de extração de características, pré-processamento, construção e avaliação dos modelos.

### 5.1 DELINEAMENTO EXPERIMENTAL E COLETA DE DADOS

O experimento foi aprovado pela CEUA - Comissão de Ética para o Uso de Animais, considerando aspectos legais e éticos das intervenções que foram realizadas, sob o protocolo nº 6957201219 (ID 001415).

O experimento foi realizado na Faculdade de Zootecnia e Engenharia de Alimentos – FZEA da Universidade de São Paulo (USP), em Pirassununga – SP, localizada a 21° 59' 46" S 47° 25' 33" W e altitude aproximada de 627 metros. O clima da região é classificado segundo Kooppen como Cwa, com duas estações bem definidas (verão chuvoso e inverno seco), com raras ocorrências de geadas. A cidade possui temperatura e pluviosidade média anual de 21,5°C e 1395 mm, respectivamente.

A coleta de dados foi realizada na Câmara Climática do Departamento de Reprodução Animal (VRA), pertencente à Faculdade de Medicina Veterinária e Zootecnia (FMVZ), da Universidade de São Paulo, durante o final do inverno e início da primavera, nos meses de setembro e outubro de 2020, totalizando 46 dias de duração.

Foram utilizadas 10 bezerras desaleitadas, da raça Holandesa, com idade aproximada de 3 a 4 meses e peso médio aproximadamente 130 Kg. Os animais tiveram acesso à água *ad libitum* e alimentação composta por feno, mistura mineral, e mistura concentrada, ajustadas semanalmente, com base em escores diários de consumo.

O experimento contou com três tratamentos buscando conhecer as respostas fisiológicas e comportamentais das bezerras em situações de temperatura de conforto e temperaturas elevadas (ondas de calor). Os tratamentos foram:

- Tratamento Tco (Temperatura de Conforto) - Ambiente de conforto (temperatura média esperada de 27°C);
- Tratamentos TEs agudo e TEs crônico (Temperatura de Estresse) - Ambiente estressante (simulando ondas de calor, com temperaturas de 30 a 35,7 °C. Para o TEs agudo foram considerados 2 dias, enquanto para o TEs crônico foram considerados 3 dias).

Os três tratamentos foram avaliados na instalação da Câmara Climática, sistema (*Tie Stall*), com uma área total de 60 m<sup>2</sup> (12 m<sup>2</sup> animal<sup>-1</sup>), com piso de concreto e 3 m de pé direito. A instalação contém 6 divisões abertas de perfil tubular, com largura de 1m e comprimento 1,80 m cada uma. Cada divisão contém um cocho de concreto com dimensões de 0,63 m de comprimento e 1 m de largura, 1 bebedouro automático individual, modelo tipo concha com disponibilidade de água contínua até atingir 3 L. Os animais ficaram amarrados em baias individuais dentro da Câmara Climática (Figura 6), podendo andar linearmente entre as divisões e tiveram acesso a todos os cochos e bebedouros. O piso da instalação foi forrado com cama de feno, numa altura aproximada de 10 cm, para facilitar o manejo sanitário do ambiente e promover conforto para descanso dos animais durante a execução do experimento.

Figura 6 – Câmara Climática



Fonte: Foto capturada pela equipe do experimento.

A Câmara Climática possui termostatos internos, exaustor e dois ventiladores, possui um controle automático de temperatura e umidade, permitindo também o controle manual pelo pesquisador, ajustando a temperatura e umidade desejadas para o estudo.

Nos tratamentos TEs agudo e TEs crônico, foram simuladas duas ondas de calor, de 5 dias consecutivos cada uma, e com um intervalo de descanso (sem estresse térmico) de 13 dias entre elas. Nas duas ondas de calor foram avaliadas as respostas fisiológicas dos animais considerando o primeiro e segundo dia de avaliação como estresse agudo, e os três dias seguintes como estresse crônico.

As ondas de calor foram simuladas com base na utilização de dados climáticos dos últimos 30 anos da região, conforme realizados por Reis et al. (2019). Buscando caracterizar como ondas de calor, episódios ocorridos de 5°C acima da temperatura média, como descrito pela Organização Meteorológica Mundial (OMM: WMO-TD n°. 1071, 2019).

Com base nos dados climáticos da região de Pirassununga-SP, entre os anos 1980 e 2013 obtidos por Xavier et al. (2015), foram realizada uma análise preliminar para verificar a média de temperatura da região nesse banco de dados. Os dados foram trabalhados juntamente com o Centro de Pesquisas Meteorológicas e Climáticas Aplicadas a Agricultura CEPAGRI- UNICAMP de Campinas-SP e, posteriormente, foi utilizado um modelo matemático, conforme Russo et al. (2014), para caracterizar as ondas de calor. As ondas de calor foram realizadas com o aumento gradual de temperatura durante os cinco dias, nos quais foram ajustadas as temperaturas às 10h para 30° C, ao 12h para 32°C e às 14h para 35,7° C, sendo reduzida gradualmente a partir das 15h. O experimento ocorreu da seguinte forma:

- *Dias 01/09 a 03/09:* Adaptação dos animais do 1° grupo. As 5 bezerras permaneceram na câmara em temperatura ambiente;
- *Dias 04/09 a 08/09:* Avaliação dos animais em temperatura de conforto;
- *Dias 09/09 a 13/09:* 1° onda de calor para o 1° grupo de animais;
- *Dia 14/09:* Intervalo de 13 dias para o primeiro grupo de animais. Neste dia os animais do 1° grupo foram transportados até o piquete do setor de gado leiteiro. O 2° grupo foi transportado até a câmara climática;

- *Dias 14/09 a 16/09*: Adaptação dos animais do 2º grupo. As 5 bezerras permaneceram na câmara em temperatura ambiente;
- *Dias 17/09 a 21/09*: Avaliação dos animais em temperatura de conforto;
- *Dias 22/09 a 26/09*: 1º onda de calor para o 2º grupo de animais;
- *Dia 27/09*: Intervalo de 13 dias para o 2º grupo de animais. Neste dia os animais do 2º grupo foram transportados até o piquete do setor de gado leiteiro. O 1º grupo foi transportado até a câmara climática;
- *Dias 28/09 a 02/10*: 2º onda de calor para o 1º grupo de animais;
- *Dia 03/10*: Devolução do 1º grupo de animais ao piquete do setor de gado leiteiro;
- *Dia 10/10*: Transporte do 2º grupo de animais até a câmara climática;
- *Dias 11/10 a 15/10*: 2º onda de calor para o 2º grupo de animais;
- *Dia 16/10*: Devolução do 2º grupo de animais ao piquete do setor de gado leiteiro.

Na câmara climática foi instalado um coletor de dados (*data logger*) modelo Onset da empresa HOBO para obtenção das variáveis ambientais temperatura do ar (T, °C) e umidade relativa (UR, %). Os equipamentos foram programados para registrar os dados ambientais a cada 15 minutos, durante todo o período experimental. E também foi instalado um termohigrômetro para o acompanhamento das temperaturas e umidade máxima e mínima dentro do ambiente.

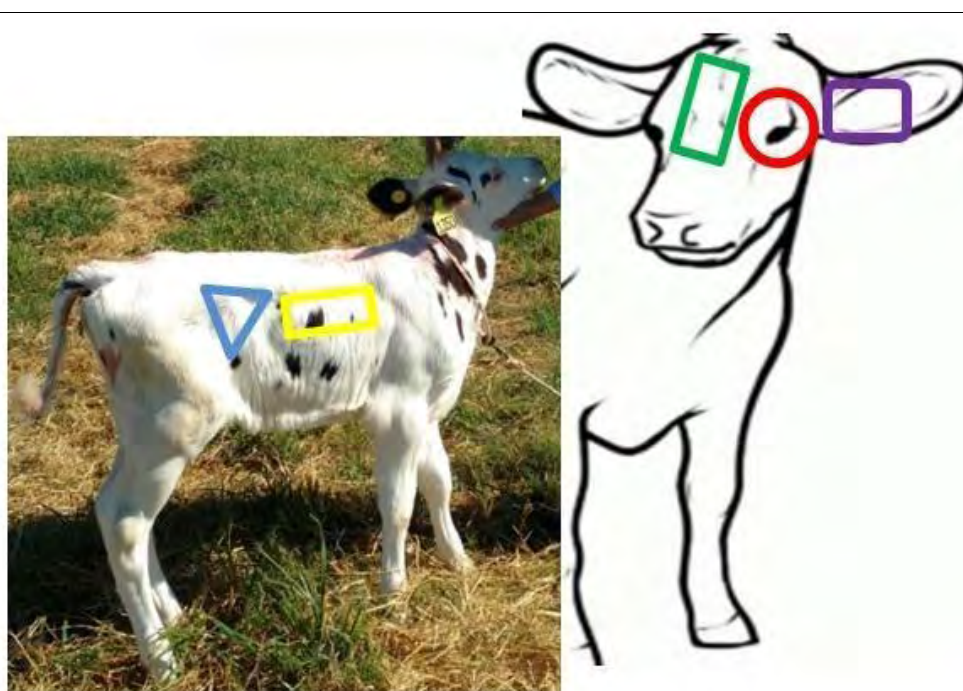
Foram registrados parâmetros fisiológicos de frequência respiratória (FR), temperatura retal (TR) e temperatura de superfície corporal por meio de TIV de todos os animais (n=10). Tais medidas foram realizadas 5 vezes ao dia (06h, 10h, 14h, 18h e 22h), nos dias de Tco e de TEs crônico, e 9 vezes ao dia (06h, 08h, 10h, 12h, 14h, 16h, 18h, 20h e 22h) nos dias de TEs agudo.

A FR foi mensurada pelo monitoramento do tempo para cada 10 movimentos do flanco, sendo calculado posteriormente o número de movimentos por minuto.

A TR foi medida diretamente na mucosa retal, por meio de um termômetro clínico digital (*TechLine*), com escala entre 20 e 50°C e precisão de  $\pm 0,1^\circ\text{C}$ .

Os dados de TIV foram registradas por meio da câmera termográfica *TESTO 875-2*, calibrada com coeficiente de emissividade de 0,98. As TIVs foram registradas de 5 regiões corporais: fronte, olho, orelha, costela e flanco (Figura 77), tomadas à distância de 1 metro do animal. Posteriormente, as TIVs coletadas foram analisadas pelo programa *TESTO IRsoft* que forneceu os perfis termográficos dos animais avaliados.

Figura 7 – Regiões da superfície corporal de onde foram coletados os dados termográficos: fronte (verde), olho (vermelho), orelha (roxo), costela (amarelo) e flanco (azul)



Fonte: Elaborado pela equipe do experimento.

## 5.2 PRÉ-PROCESSAMENTO E SELEÇÃO DOS DADOS

Os dados de termografia foram organizados em uma base de dados e os outros dados coletados (ambientais e fisiológicos) foram organizados em uma planilha, totalizando 910 registros contendo dados ambientais e fisiológicos, cada um deles contendo 3 imagens termográficas.

Na etapa de pré-processamento foi realizada a detecção de dados com valores discrepantes (*outliers*) de TR e FR. O procedimento foi feito utilizando

dois diferentes métodos de detecção de *outliers*: o algoritmo de Tukey (*Tukey's fences*) e o Z-Score.

O algoritmo de Tukey consiste em definir um limite superior e um limite inferior a partir da distância interquartil dos dados. Os limites são calculados com as eq. (1) e (2).

$$LS = Q3 + K(Q3 - Q1) \quad (1)$$

$$LI = Q1 - K(Q3 - Q1) \quad (2)$$

Em que:

$LS$  é o limite superior;

$LI$  é o limite inferior;

$Q3$  é o 3º quartil;

$Q1$  é o 1º quartil.

Nas Equações 1 e 2 o parâmetro  $K$  representa a tolerância da detecção de *outliers*, para o qual normalmente se utiliza o valor de 1,5. Um dado é considerado *outlier* se for maior que  $LS$  ou menor que  $LI$ .

O método Z-Score consiste em avaliar a quantos desvios padrão um dado está da média, e é calculado com a eq. (3):

$$Z_n = \frac{x_n - \bar{x}}{\sigma_x} \quad (3)$$

Em que:

$Z_n$  é o Z-Score do  $n$ -ésimo dado;

$x_n$  é o valor do  $n$ -ésimo dado;

$\bar{x}$  é a média dos dados;

$\sigma_x$  é o desvio padrão dos dados.

O valor  $Z_n$  indica quantos desvios padrão o valor do dado  $n$  difere da média e usualmente adota-se  $Z_n = 3$  como limiar para definir o dado como *outlier*. Além disso, o método Z-Score é também útil para avaliar em termos quantitativos o quanto um dado é *outlier*.

Uma vez escolhidos os métodos, estes foram implementados para detecção dos *outliers* de TR e de FR da base de dados. Para o algoritmo de

Tukey foram escolhidos valores de 1, de 1,3 e de 1,5 para o parâmetro de tolerância  $K$ . O método Z-Score foi utilizado para fins de análise e para a validação do algoritmo de Tukey. Para avaliar o impacto da exclusão dos outliers no desempenho da classificação foram construídos modelos classificadores a serem detalhados nas próximas seções.

Após a etapa da detecção de *outliers*, foi feita uma seleção de dados a serem usados no treinamento dos modelos classificadores. Tendo uma base de dados composta por imagens termográficas de 5 diferentes regiões do corpo dos animais, este trabalho buscou avaliar o potencial de cada uma das regiões como preditoras do estresse térmico (seleção de atributos de TIV). A comparação entre as regiões foi feita a partir da análise de correlação, sendo calculado o coeficiente de correlação de Pearson das temperaturas mínima, média e máxima encontradas nas matrizes de temperatura de cada região com as variáveis fisiológicas TR e FR, que são comumente associadas ao estresse. Os dados referentes às duas regiões que apresentaram as maiores correlações foram selecionados para serem usados no treinamento dos modelos.

Além da análise de correlação dos dados completos, foram calculados também os mesmos coeficientes de correlação para a base de dados separada em registros referentes aos dias em que os animais estiveram submetidos à 1ª onda de calor e os referentes à 2ª onda de calor (seleção de registros). A finalidade desta análise é verificar se a resposta termorregulatória dos animais ao ambiente de estresse foi diferente de uma onda para a outra, para que se possa então otimizar os modelos de acordo com eventuais diferenças.

A combinação dos procedimentos de seleção de atributos de TIV e seleção de registros apresentados nos dois últimos parágrafos foi usada para estruturar diferentes subconjuntos de dados a serem usados como entradas para os modelos na fase de treinamento, juntamente com as variáveis meteorológicas T e UR, a fim de comparar os resultados obtidos por cada um deles.

### 5.3 EXTRAÇÃO DE CARACTERÍSTICAS

Para que as imagens termográficas dos animais pudessem ser classificadas por meio de algoritmos de aprendizado de máquina, foi definido inicialmente quais atributos preditores seriam usados como entradas dos

modelos. Como o objetivo deste trabalho é classificar o estresse térmico a partir de imagens termográficas e dados ambientais, foi necessário implementar um método de extração de características para que fossem extraídos atributos preditores das imagens, a serem utilizados pelos modelos juntamente com dados ambientais (temperatura do ar e umidade relativa) para classificação do estresse térmico. O método de extração de características adotado neste trabalho é denominado Assinatura Térmica e está descrito nesta seção (RODRIGUES, 2022).

Inicialmente, as imagens coletadas das diferentes partes do corpo do animal foram recortadas em imagens menores de 24 x 24 pixels, sendo posteriormente exportadas como matrizes de valores de temperatura (formato de arquivo xlsx), onde cada célula da matriz possui um valor de temperatura correspondente a um pixel da respectiva imagem termográfica.

A assinatura térmica é um tipo de histograma que mostra a frequência relativa de temperaturas de cada imagem dentro de uma faixa de temperatura pré-estabelecida. Sendo assim, a assinatura térmica é um atributo descritor que contém a distribuição percentual das células de uma matriz de temperaturas em cada faixa.

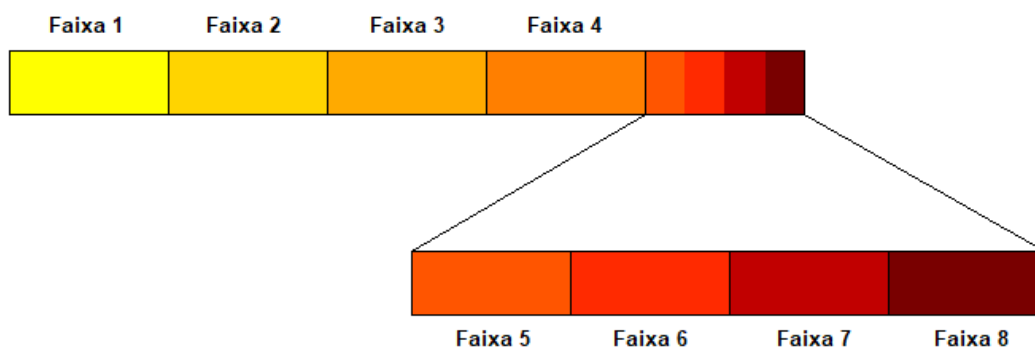
Para que seja determinada a assinatura térmica de cada matriz de temperaturas, é preciso primeiro estabelecer os intervalos de temperatura de cada faixa em relação à toda a base de dados de termografia. A delimitação dos intervalos deve ser feita de tal modo que as faixas de temperatura sejam balanceadas (distribuição mais homogênea de temperaturas entre os intervalos) e representativas da base de dados de termografia como um todo. Além disso, é importante considerar que como o objetivo deste método gerar um atributo descritor do estresse pelo calor, as faixas mais quentes de temperatura têm um maior potencial como descritoras do estresse térmico do que as faixas mais frias, sendo necessário que a assinatura térmica tenha uma maior precisão na avaliação de temperaturas mais altas. Para lidar com estes problemas, foi utilizado o método dos quantis.

Inicialmente, foi contabilizada para cada uma das 5 regiões a soma do número total de células de todas as matrizes de temperatura. As células são arranjadas em ordem crescente de temperaturas para que sejam calculados os quantis. Primeiramente, dividiu-se o total de células em 5 quantis e em seguida



o último quantil foi dividido em 4 quantis, de modo a obter-se uma assinatura térmica de 8 faixas, onde as 4 faixas mais frias são mais largas e as 4 faixas mais quentes são mais estreitas, garantindo assim uma maior precisão na avaliação de altas temperaturas. O procedimento está ilustrado na Figura 8.

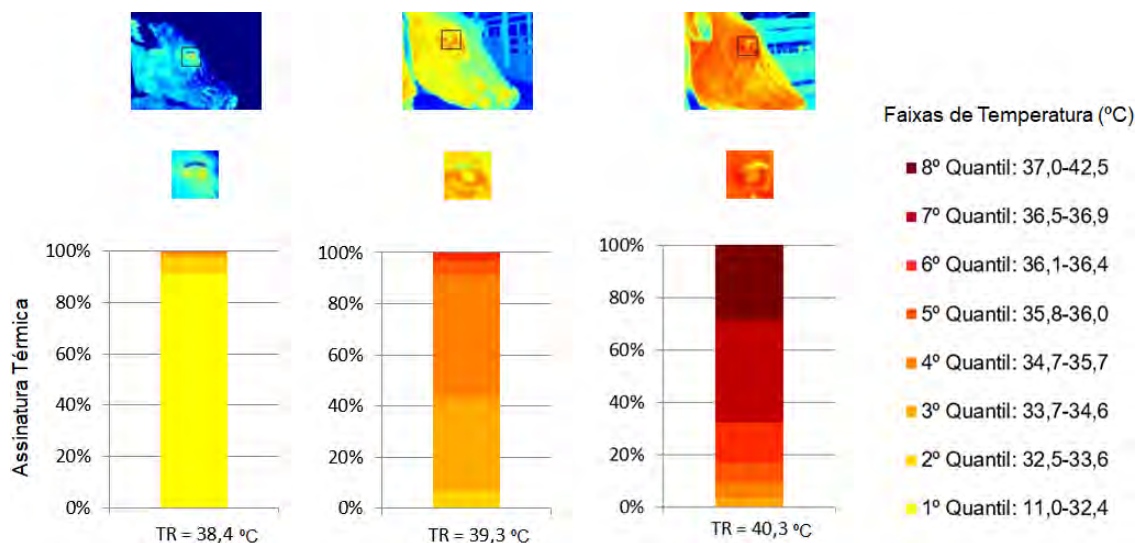
Figura 8 – Delimitação das faixas da assinatura térmica



Fonte: própria autoria.

Uma vez definido o método para delimitação, foram determinados os intervalos de temperatura para a partir do total de células das matrizes de temperatura. A Figura 9 mostra 3 exemplos de imagens termográficas da região do olho com os respectivos recortes menores e assinaturas térmicas, bem como os intervalos das faixas de assinatura térmica determinados para a região do olho. Pode-se observar que a assinatura térmica referente à imagem do animal com TR de 40,3 °C contém uma melhor distribuição de temperaturas entre as faixas, o que mostra que o método de delimitação das faixas foi efetivo em descrever temperaturas altas com maior precisão.

Figura 9 – Exemplos de assinaturas térmicas obtidas para diferentes condições de estresse térmico definidas pela temperatura retal



Fonte: própria autoria.

Após a delimitação das faixas de temperatura, a assinatura térmica de cada matriz de temperatura foi obtida e armazenada para posterior treinamento dos modelos de aprendizado de máquina.

A avaliação permite identificar de qual parte do corpo do animal pode a termografia gerar o melhor descritor.

#### 5.4 CONSTRUÇÃO DOS MODELOS

Para a construção dos modelos classificadores, é preciso primeiramente fazer a rotulagem dos dados, de modo que uma vez rotulados em classes de estresse térmico eles possam ser utilizados por algoritmos de aprendizado de máquina para treinamento dos modelos. Usualmente, o estresse térmico em animais de produção é avaliado em termos da TR ou da FR, sendo definidas faixas de valores destas variáveis fisiológicas para a classificação do estresse.

Para organização do banco de dados para modelagem foram feitos 3 diferentes tipos de rotulagem de dados, sendo que dois deles têm como referência a FR e um deles tem como referência a TR. Quanto às rotulagens baseadas na FR, uma delas é feita em 2 classes (Estresse e Não-estresse) e a outra em 3 classes (Conforto, Alerta e Perigo). A rotulagem baseada em TR é

feita em 2 classes (Estresse e Não-estresse). Neste trabalho os tipos de rotulagem de dados são denotados como:

- 2C/TR: rotulagem baseada em TR e feita em 2 classes;
- 2C/FR: rotulagem baseada em FR e feita em 2 classes;
- 3C/FR: rotulagem baseada em FR e feita em 3 classes.

Com isso é possível criar modelos para predizer o nível de estresse térmico em 2 classes e em 3 classes de acordo com o conjunto de entradas apresentado. Os modelos de 2 classes são sugeridos como adequados para uma maior praticidade na classificação, enquanto o de 3 classes é adequado para um maior detalhamento no reconhecimento dos níveis estresse dos animais.

As faixas de valores de TR e FR definidoras dos níveis de estresse térmico para cada tipo de rotulagem de dados estão descritas nas Tabelas 1 e 2.

Tabela 1 – Níveis de estresse para os tipos de rotulagem 2C/FR e 2C/TR

Nível de estresse térmico	2C/FR (mov/min)	2C/TR (°C)
Não-estresse	FR < 40	TR ≤ 39,3
Estresse	FR ≥ 40	TR > 39,3

Fonte: ROBINSON E.N. **Termorregulação**. In: CUNNINGHAM, J.G. Tratado de fisiologia veterinária. 2.ed. Rio de Janeiro: Guanabara Koogan, cap.51, 427-435, 1999.

Legenda: 2C/FR – rotulagem baseada em FR e feita em 2 classes; 2C/TR – rotulagem baseada em TR e feita em 2 classes

Tabela 2 – Níveis de estresse para o tipo de rotulagem 3C/FR

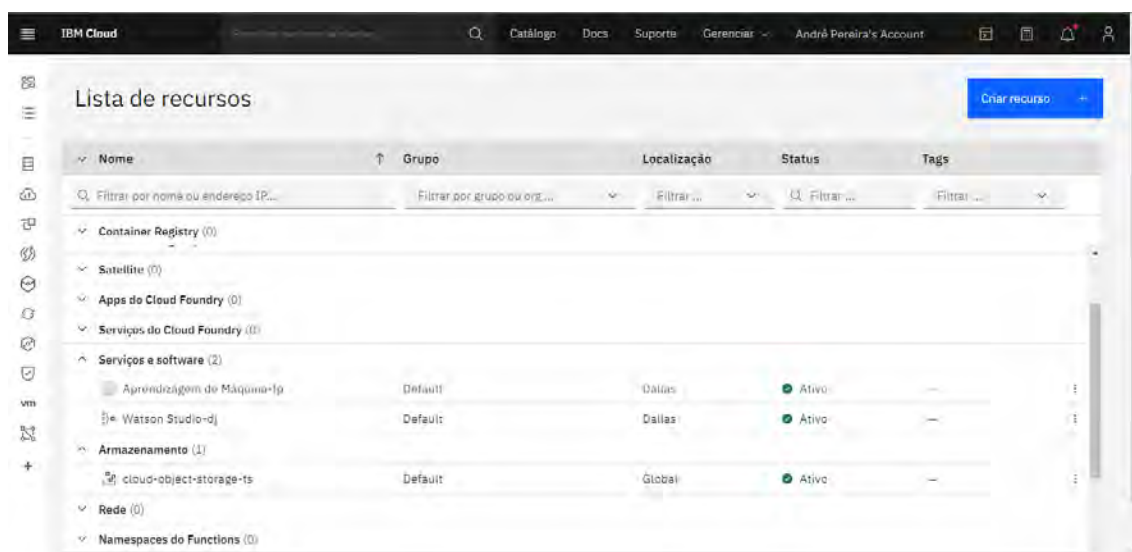
Nível de estresse térmico	3C/FR (mov/min)
Conforto	FR < 40
Alerta	40 ≤ FR ≤ 60
Perigo	FR > 60

Fonte: ROBINSON E.N. **Termorregulação**. In: CUNNINGHAM, J.G. Tratado de fisiologia veterinária. 2.ed. Rio de Janeiro: Guanabara Koogan, cap.51, 427-435, 1999.

Legenda: 3C/FR – rotulagem baseada em FR e feita em 3 classes

Uma vez feita a rotulagem dos dados, foi utilizada a plataforma da IBM para a construção dos modelos. Para utilizar os recursos disponibilizados pela IBM, foi criada uma conta na IBM Cloud, o serviço de armazenamento em nuvem da IBM. A partir do catálogo de recursos foram adicionados os recursos Watson Studio e Aprendizado de Máquina. A interface da IBM Cloud com a lista de recursos adicionados está ilustrada na Figura 10.

Figura 10 – Captura de tela da interface da IBM Cloud



Fonte: própria autoria.

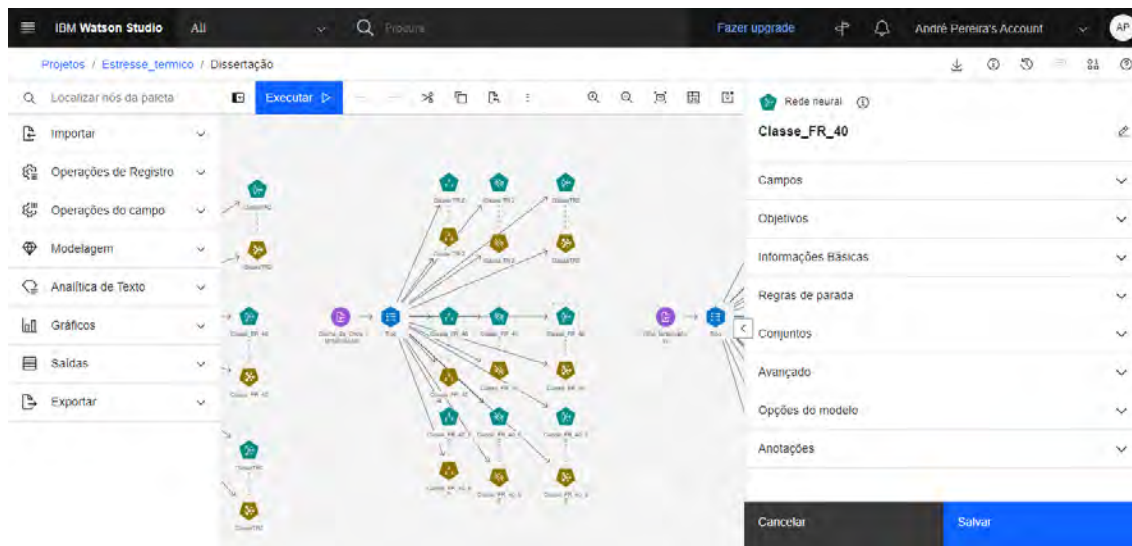
Uma vez aberto o Watson Studio, foi criado um projeto para realizar o desenvolvimento dos diferentes modelos a serem avaliados. Foi utilizada a base de dados organizada contendo os dados referentes às variáveis ambientais, à assinatura térmica e às classes de estresse térmico.

Inicialmente foi criado um modelador do tipo SPSS Modeler, que é um ambiente voltado para mineração de dados e criação de modelos de aprendizado de máquina. O SPSS Modeler funciona como um fluxograma, onde o fluxo de dados passa por blocos nos quais são realizadas operações de diversos tipos, tais como operações sobre os registros, operações sobre os atributos, modelagem de classificação e predição e construção de gráficos para representação dos dados.

A Figura 11 ilustra um exemplo da interface do SPSS Modeler contendo a estrutura de blocos representando o fluxo e processamento dos dados, bem como o menu de opções de blocos para realizar diferentes operações sobre os

dados (menu à esquerda) e o menu de configurações do bloco de modelagem com algoritmo de RNA (menu à direita).

Figura 11 – Captura de tela da interface do SPSS Modeler



Fonte: própria autoria.

Um dos blocos de modelagem é o bloco Auto Classifier, que tem como função a construção automática de modelos a partir de uma lista de algoritmos de classificação disponíveis, sem que seja necessário que o usuário escolha quais algoritmos e parâmetros utilizar. Ao ser conectado ao fluxo de dados e executado, ele constrói modelos utilizando os algoritmos selecionados da lista e retorna os modelos de maior acurácia, indicando o algoritmo utilizado para a construção de cada um deles. O Auto Classifier é útil como uma forma de identificar os algoritmos que melhor classificam os dados dentre os disponibilizados pelo SPSS Modeler, sendo, portanto, utilizado neste trabalho com esta finalidade.

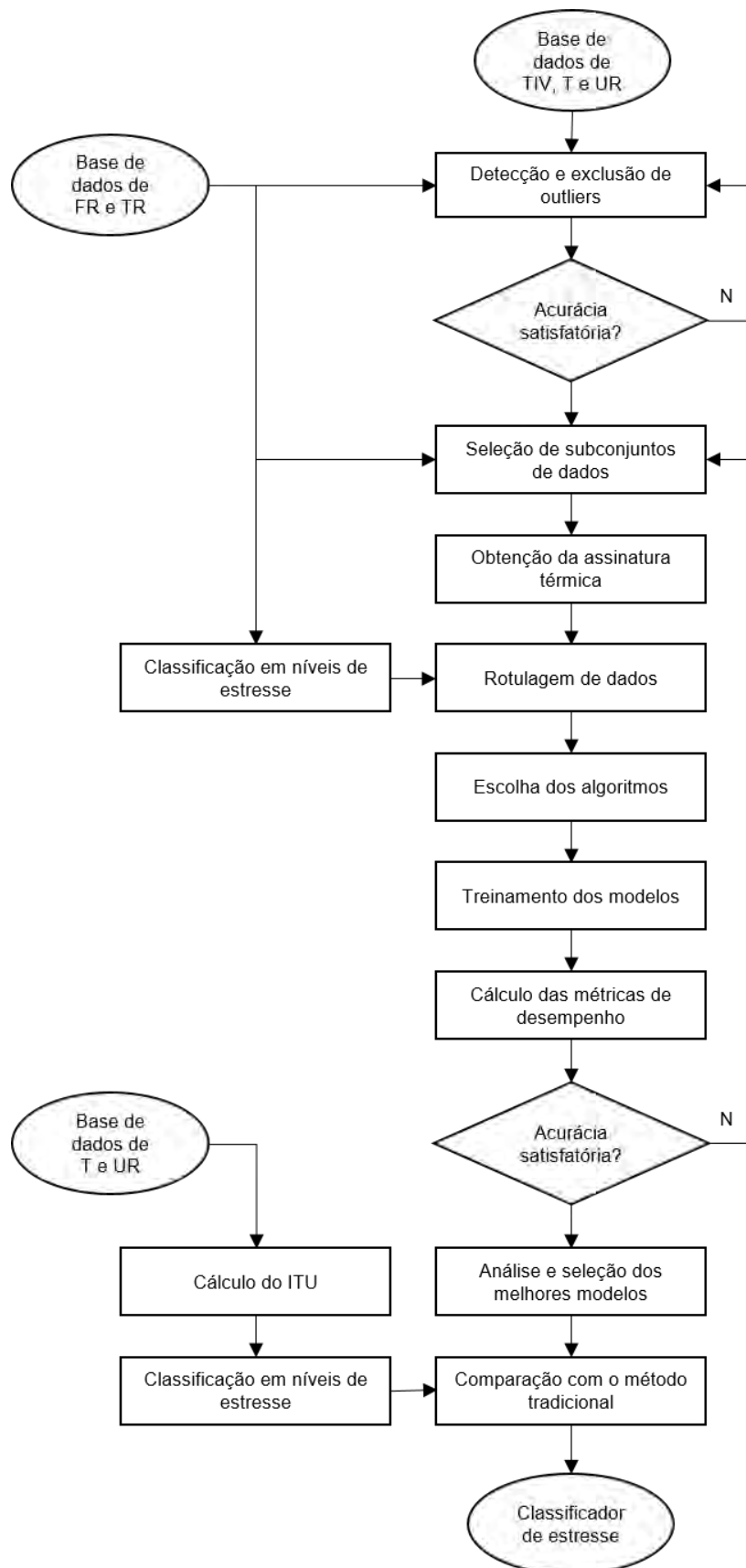
Embora o Auto Classifier seja útil para determinar quais algoritmos melhor realizam a tarefa de classificação, ele não consegue aproveitar ao máximo o potencial de cada algoritmo, visto que não possui a opção de ajustar os parâmetros de cada um individualmente. Portanto, é necessário que cada algoritmo identificado pelo Auto Classifier seja em seguida aplicado individualmente. O procedimento adotado neste trabalho foi então utilizar numa primeira etapa o Auto Classifier para identificar os três melhores algoritmos para

a classificação (pré-seleção) e, posteriormente, aplicar os algoritmos separadamente para obtenção de melhores modelos.

Como uma forma de sintetizar as etapas de construção dos modelos classificadores de estresse e a análise dos resultados obtidos, apresenta-se na Figura 12 o processo de descoberta de conhecimento inspirado no KDD. Os passos 1 a 8 compõe o processo de construção dos modelos enquanto os passos 9 e 10 compõe a análise dos resultados obtidos, a qual está detalhada na próxima seção.

1. Obtenção e organização da base de dados;
2. Detecção e exclusão de *outliers*;
3. Seleção de subconjuntos de dados (seleção de atributos de TIV e seleção de registros) com maior potencial de predição do estresse;
4. Obtenção da assinatura térmica;
5. Rotulagem dos dados de acordo com diferentes critérios;
6. Escolha dos algoritmos para treinamento dos modelos através do bloco Auto Classifier;
7. Treinamento dos modelos;
8. Cálculo das métricas de desempenho (a ser detalhado na próxima seção);
9. Análise e seleção dos melhores modelos;
10. Comparação com o método tradicional.

Figura 12 – Fluxograma de construção dos modelos e análise dos resultados



Fonte: própria autoria.

Legenda: FR – frequência respiratória; TR – temperatura retal; TIV – imagens termográficas.

Na Figura 12, as setas ao longo dos blocos centrais indicam o fluxo de dados pelo processo, os blocos em formato de elipse indicam diferentes partes da base de dados e à direita estão representados os ciclos de realimentação de dados. Por se tratar de uma abordagem não-trivial, a formulação de um processo de descoberta de conhecimento envolve uma série de retornos no fluxo de dados para que os detalhes do processo sejam ajustados em função do desempenho obtido pelos modelos. Os ciclos de realimentação indicam os blocos que tiveram algum tipo de ajuste de parâmetro ou procedimento feito a partir de testes preliminares para otimizar o desempenho do processo como um todo. Dois ciclos de realimentação foram adotados neste trabalho, um para a etapa “Detecção e Exclusão de *Outliers*” e outro para as etapas “Seleção de Subconjuntos de Dados”, “Rotulagem de Dados” e “Escolha do Algoritmo”.

Na etapa “Detecção e Exclusão de *Outliers*” houve o ajuste do parâmetro de tolerância  $K$  conforme resultados de testes preliminares. Os testes foram feitos através da escolha de diferentes valores de  $K$  do algoritmo de Tukey para detecção de *outliers* de FR e de TR e posterior geração de modelos classificadores de estresse. Os modelos foram gerados no SPSS Modeler utilizando-se o algoritmo de RNA, com dados rotulados de acordo com o tipo de rotulagem 2C/FR (para detecção de *outliers* de FR) e o 2C/TR (para detecção de *outliers* de TR). A acurácia dos modelos foi utilizada como referência para avaliar o impacto que a exclusão de *outliers* teve no processo e o parâmetro  $K$  foi ajustado até que a acurácia dos modelos convergisse para valores satisfatórios.

Nas etapas “Seleção de Subconjuntos de Dados”, “Rotulagem de Dados” e “Escolha do Algoritmo” foram realizados diversos testes preliminares a fim de se definir, respectivamente, quais critérios para seleção dos dados, quais tipos de rotulagem e quantos algoritmos de classificação diferentes seriam avaliados neste trabalho. Após os testes preliminares foram definidos então três parâmetros, aqui denominados Parâmetros Primários de Modelagem:

- Subconjunto de Dados: resultado da seleção de atributos de TIV e seleção de registros;
- Tipos de Rotulagem de Dados: 2C/FR, 3C/FR e 2C/TR;



- Algoritmo de Classificação: pré-seleção dos três melhores algoritmos pelo Auto Classifier.

Os modelos finais foram criados a partir das diferentes combinações dos três parâmetros primários. Dessa forma, com o processo proposto busca-se ajustar os parâmetros primários de modo a otimizar o desempenho dos modelos.

Na próxima seção descreve-se como os modelos finais resultantes deste processo de modelagem foram avaliados e comparados entre si e com o método tradicional de avaliação de estresse baseado no ITU.

## 5.5 AVALIAÇÃO DOS MODELOS E FORMA DE ANÁLISE DOS RESULTADOS

Uma vez que os modelos tenham sido desenvolvidos, é necessário avaliá-los a fim de determinar qual possui o melhor desempenho, procedimento este que é feito da seguinte forma. Antes da criação dos modelos, é separada a base de dados em duas partes, sendo o conjunto de dados para treinamento utilizado pelos algoritmos na construção dos modelos, e o conjunto de dados para teste utilizado na comparação entre modelos. Esta separação é necessária para que os modelos não sejam testados com dados já conhecidos, o que prejudica a avaliação. Uma vez prontos, os modelos foram usados para classificar os dados de teste, e a classe predita pelo modelo é comparada com o a classe real definida na rotulagem dos dados.

Idealmente, se houvesse dados em quantidade grande o suficiente, a separação em conjuntos de treinamento e de teste seria fixa (procedimento denominado *hold out*), mas como na maioria das aplicações os dados são escassos este não é um método adequado. Para contornar este problema, utiliza-se o método da validação cruzada *K-fold* (*K-fold cross-validation*). Neste método, a base de dados é dividida em um número  $K$  de partes (*folds*) aproximadamente iguais, sendo que uma delas é selecionada para ser o conjunto de teste enquanto as outras são utilizadas como conjunto de treinamento para construir o modelo. O processo se repete para cada uma das  $K$  partes, o que resulta na construção de  $K$  modelos distintos, que são testados cada um com seu respectivo conjunto de teste. É calculada então a média das métricas de desempenho obtidas por cada modelo, que será considerada a

medida de desempenho geral do modelo (HASTIE; TIBSHIRANI; FRIEDMAN, 2017).

O valor de  $K$  adotado costuma ser de 5 a 10 na maioria das aplicações. Entretanto, pode-se também adotar um valor de  $K$  igual ao tamanho da base de dados, ou seja, separar 1 dado apenas como conjunto de teste, procedimento conhecido como *leave-one-out*. Hastie, Tibshirani e Friedman (2017) afirmam que a escolha de  $K$  é uma decisão de *trade-off* entre o enviesamento e a variância da estimativa do erro do modelo, sendo que  $K$  igual a 5 ou 10 evita níveis excessivos tanto de enviesamento quanto de variância. Neste trabalho foi adotado  $K = 10$ .

Após passar os dados de teste pelos modelos, foram calculadas as métricas de desempenho. Estas métricas são calculadas de maneira distinta para modelos de classificação e modelos de predição.

Os modelos de predição normalmente são avaliados através de análise de regressão entre a variável medida e a variável predita. Por outro lado, os modelos de classificação são avaliados através de uma ferramenta de análise chamada Matriz de Confusão. Em uma tarefa de classificação binária (com apenas duas classes, por exemplo Sim e Não), os resultados obtidos por um modelo podem ser divididos em:

- Verdadeiros positivos (VP), quando o dado foi acertadamente classificado como Sim;
- Verdadeiro negativo (VN), quando foi acertadamente classificado como Não;
- Falso positivo (FP), quando foi erroneamente classificado como Sim;
- Falso negativo (FN), quando foi erroneamente classificado como Não.

Ao passar os dados de teste pelo modelo, os resultados podem ser organizados em uma matriz como exemplificado pela Tabela 3.

Tabela 3 – Modelo de uma matriz de confusão

Valor real observado	Valor predito pelo modelo	
	Sim	Não
Sim	VP	FN
Não	FP	VN

Fonte: Própria autoria

A partir dos resultados de VP, VN, FP e FN, foram calculadas diversas métricas de desempenho (VIJAY; BALA, 2015):

- Precisão, que é a proporção de casos selecionados pelo modelo que são de fato relevantes, podendo ser também entendida como a capacidade de evitar resultados falsos positivos. É calculada como  $VP/(VP+FP)$ ;
- Sensibilidade, que é a habilidade do modelo de selecionar todos os casos que devem ser de fato selecionados, podendo ser também entendida como a capacidade de evitar resultados falsos negativos. É calculada como  $VP/(VP+FN)$ ;
- F1 Score, que é a combinação das duas últimas métricas em uma. Definida como a média harmônica entre a precisão e a sensibilidade, é calculada como  $2*precisão*sensibilidade/(precisão+sensibilidade)$ ;
- Acurácia, que é a habilidade geral do modelo de obter bons resultados, calculada como  $VP+VN/(VP+VN+FP+FN)$ .

A depender de como se deseja avaliar o modelo e qual a aplicação, é mais apropriado analisar uma ou outra métrica. A acurácia é uma medida geral de desempenho, mas ela pode ser inadequada caso os dados a serem classificados sejam desbalanceados, o que torna o F1 Score uma medida preferível. Algumas aplicações exigem que se tenha uma tolerância menor com erros do tipo falso positivo ou com falso negativos, o que torna a precisão ou a sensibilidade, respectivamente, mais prioritárias (VIJAY; BALA, 2015).

Caso o problema seja de classificação multiclasse (com mais de 2 classes), as métricas são calculadas individualmente para cada classe e em seguida é feita a média entre elas (SOKOLOVA; LAPALME, 2009).

Os modelos classificadores gerados pelo SPSS Modeler tiveram suas acurácias calculadas diretamente pelos blocos dedicados à construção deles, através do método de validação cruzada.

Após a construção dos modelos e obtenção das métricas de desempenho, foi calculada a acurácia normalizada deles. Como entre os modelos gerados existem modelos que classificam em 2 classes e em 3 classes, não é possível fazer uma comparação direta entre eles. Deve-se levar em conta que se for feita uma classificação aleatória em 2 classes, como em um lançamento de uma moeda, a acurácia será de aproximadamente 50% (probabilidade de acerto), enquanto uma classificação aleatória em 3 classes terá acurácia de aproximadamente 33% (probabilidade de acerto). A normalização da acurácia corrige essa diferença e permite comparar diretamente modelos que classificam em números diferentes de classes. O cálculo da acurácia normalizada foi feito com a eq. (4).

$$AN_n = \frac{A_n - x}{1 - x} \quad (4)$$

Em que:

$AN_n$  é a acurácia normalizada do modelo  $n$ ;

$A_n$  é a acurácia do modelo  $n$ ;

$x$  é igual a  $1/3$  se o modelo classifica em 3 classes e  $1/2$  se o modelo classifica em 2 classes.

Foram selecionados os modelos de maior acurácia dentre os que classificam em 2 classes e dentre os que classificam em 3 classes, e foram obtidas as matrizes de confusão de cada um. Uma vez obtidas as matrizes de confusão dos modelos, foi comparado o desempenho deles com o método tradicional de avaliação do estresse, feito a partir de índices de conforto térmico. Para tanto, calculou-se o ITU de acordo com a eq. (5) a partir do modelo definido por Thom (1959):

$$ITU = 0,8 \times T_{ar} + \left(\frac{UR}{100}\right) \times (T_{ar} - 14,4) + 46,4 \quad (5)$$

Em que:

$T_{ar}$  = temperatura do ar (°C);

$UR$  = umidade relativa do ar (%).

Em seguida, utilizou-se o critério de Rosenberg et al. (1983) para classificar o estresse térmico a partir do ITU. Originalmente, o critério de Rosenberg et al. (1983) classifica o estresse em 4 classes distintas, sendo necessário fazer uma adaptação para este trabalho, que adota classificações em 2 e 3 classes. As classes de estresse com base no ITU estão definidas nas Tabelas 4 e 5.

Tabela 4 – Faixas de valores de ITU para classificação de estresse térmico em 2 classes

<b>Classe de estresse térmico</b>	<b>Faixas de valores de ITU</b>
Não Estresse	ITU < 75
Estresse	ITU ≥ 75

Fonte: adaptado de Rosenberg et al. (1983)

Tabela 5 – Faixas de valores de ITU para classificação de estresse térmico em 3 classes

<b>Classe de estresse térmico</b>	<b>Faixas de valores de ITU</b>
Conforto	ITU < 75
Alerta	75 ≤ ITU < 79
Perigo	ITU ≥ 79

Fonte: adaptado de Rosenberg et al. (1983)

Após a classificação dos dados com base no ITU, foram geradas as matrizes de confusão com as métricas de desempenho, as quais representam o desempenho do método tradicional a ser comparado com o método proposto neste trabalho.

A análise dos resultados foi feita em duas formas descritas a seguir:

- Análise 1: comparação dos modelos de aprendizado de máquina. Foram gerados modelos a partir das diferentes combinações dos três parâmetros primários (subconjunto de dados para treinamento, tipo de rotulagem de dados e algoritmo de classificação utilizado) e foram comparadas as acurácias e acurácias normalizadas deles.
- Análise 2: comparação dos modelos de aprendizado de máquina com o método tradicional pelo ITU. Foram selecionados da Análise 1 os modelos de maior acurácia de 2 classes e de 3 classes e suas matrizes de confusão foram comparadas com as obtidas pelo método do ITU.

Na Análise 1 buscou-se determinar como os 3 parâmetros primários devem ser ajustados para obtenção de melhores modelos, ou seja, qual o melhor subconjunto de dados, o melhor tipo de rotulagem e o melhor algoritmo. Na Análise 2 buscou-se comparar os melhores modelos selecionados na Análise 1 com os modelos construídos a partir do ITU a fim de verificar a utilidade do método proposto neste trabalho para classificação de estresse térmico.

## 6 RESULTADOS E DISCUSSÃO

Inicialmente, são apresentados os resultados obtidos pelo método de detecção de *outliers*. Em seguida a estatística descritiva dos dados e a análise de correlação. Depois os resultados da modelagem com as comparações dos modelos obtidos. Por fim a comparação dos melhores modelos com o método tradicional.

### 6.1 DETECÇÃO DE *OUTLIERS*

Após os testes preliminares, decidiu-se utilizar valores de 1,5, 1,3, e 1 para o parâmetro de tolerância  $K$  a fim de avaliar o impacto da exclusão de *outliers* no desempenho da classificação. O número de *outliers* detectados pelo algoritmo de Tukey para cada valor do parâmetro de tolerância  $K$  está apresentado na Tabela 6.

Tabela 6 – Número de *outliers* detectados para cada valor do parâmetro de tolerância  $K$  e cada variável.

Parâmetro de tolerância $K$	Variável	
	TR	FR
1	138	90
1,3	78	27
1,5	54	21

Fonte: própria autoria

Legenda: TR – temperatura retal; FR – frequência respiratória.

Durante a análise dos resultados da detecção de *outliers*, notou-se que um dos animais foi consistentemente detectado como *outlier* em um período de 3 dias do experimento. Isso se deve ao fato que este animal esteve com febre no período, o que ao ser detectado pelo método evidenciou a validade deste.

A acurácia dos modelos de teste com os diferentes valores do parâmetro  $K$  estão apresentadas na Tabela 7.

Tabela 7 – Acurácia dos modelos de teste para cada valor do parâmetro de tolerância  $K$  e base de dados sem exclusão de *outliers* e para cada variável

Parâmetro de tolerância $K$	Variável	
	TR	FR
1	70,6%	73,5%
1,3	70,6%	84,9%
1,5	70,3%	85,4%
Sem exclusão de outliers	71,7%	85,1%

Fonte: própria autoria

Legenda: TR – temperatura retal; FR – frequência respiratória.

Como pode-se observar, os valores de acurácia não apresentaram aumento com a exclusão de *outliers* da base de dados, ao contrário do que era esperado. Em razão disso, decidiu-se por utilizar a base de dados completa, sem exclusão de *outliers*, para a construção dos modelos classificadores.

## 6.2 ESTATÍSTICA DESCRITIVA, ANÁLISE DE CORRELAÇÃO E SELEÇÃO DE DADOS

A estatística descritiva dos dados fisiológicos e meteorológicos obtidos no experimento está apresentada na Tabela 8. Observa-se que foi possível obter dados dos animais em condição de estresse térmico, considerando-se tanto a FR como a TR.

Tabela 8 – Estatística descritiva dos dados fisiológicos e meteorológicos

Variável	Média $\pm$ Desvio Padrão	Mínimo	Continua
			Máximo
FR (mov/min)	58,4 $\pm$ 19	23,4	151,5
TR (°C)	39,2 $\pm$ 0,5	37,7	41,0
TIVOI (°C)	Max	37,1 $\pm$ 0,8	39,4
	Max	34,4 $\pm$ 1,3	37,4
TIVFron (°C)	Med	32,6 $\pm$ 1,8	36,5
	Min	30,7 $\pm$ 2,4	36,0



Variável		Média ± Desvio Padrão	Conclusão	
			Mínimo	Máximo
TIV <sub>Ore</sub> (°C)	Max	37,1 ± 1,2	27,3	39,8
	Med	36,2 ± 1,3	26,7	38,8
	Min	34,9 ± 1,7	25,8	38,3
TIV <sub>Cos</sub> (°C)	Max	34,3 ± 1,6	29,5	39,0
	Med	33,2 ± 1,7	28,2	38,1
	Min	32 ± 2	25,4	37,4
TIV <sub>Flan</sub> (°C)	Max	34,6 ± 1,5	30,1	38,6
	Med	33,5 ± 1,6	27,6	37,8
	Min	32 ± 1,9	24,0	36,9
T (°C)		28,5 ± 3,4	21,6	35,9
UR (%)		52,4 ± 10,7	31,3	80,3
Entapia (kJ/kg)		63,4 ± 7,3	47,1	84,2

Fonte: própria autoria.

Legenda: Dados meteorológicos: T – temperatura média do ar; UR – umidade relativa do ar. Dados fisiológicos: FR – frequência respiratória; TR – temperatura retal; TIV<sub>Fron</sub> – temperatura de superfície da fronte dos animais obtidas com termografia; TIV<sub>OI</sub> – temperatura de superfície da região ocular dos animais obtidas com termografia; TIV<sub>Ore</sub> – temperatura de superfície da orelha dos animais obtidas com termografia; TIV<sub>Flan</sub> – temperatura de superfície do flanco dos animais obtidas com termografia; TIV<sub>Cos</sub> – temperatura de superfície da costela dos animais obtidas com termografia. Parâmetros: Méd – média; Min – mínima; Max – máxima.

As Tabelas 9 e 10 apresentam o número e a porcentagem de registros da base de dados separados por nível de estresse térmico, de acordo com os três tipos de rotulagem de dados definidos na Seção 4.5. Os animais estiveram submetidos a condições ambientais majoritariamente estressantes, como observado pelo grande número de casos de FR elevada, mas que a resposta termorregulatória dos animais foi efetiva em manter a homeostase na maioria das vezes, como visto pela baixa porcentagem de casos de TR elevada.

Tabela 9 – Número e porcentagem de registros por nível de estresse, 2 classes

Nível de estresse	2C/FR		2C/TR	
	Número	Porcentagem	Número	Porcentagem
Não-estresse	435	15,9%	1476	54,1%
Estresse	2295	84,1%	1254	45,9%

Fonte: própria autoria

Legenda: 2C/FR – rotulagem de dados baseada em frequência respiratória e feita em 2 classes; 2C/TR – rotulagem de dados baseada em temperatura retal e feita em 2 classes.

Tabela 10 – Número e porcentagem de registros separados por nível de estresse, 3 classes

Nível de estresse	3C/FR	
	Número	Porcentagem
Conforto	276	15,9%
Alerta	495	28,4%
Perigo	969	55,7%

Fonte: própria autoria

Legenda: 3C/FR – rotulagem de dados baseada em frequência respiratória e feita em 3 classes

Os coeficientes de correlação de Pearson relacionando as variáveis fisiológicas estão representados na Tabela 11.

Tabela 11 – Coeficiente de Correlação de Pearson relacionando variáveis fisiológicas

Variável	FR	TR	TIVOI			TIVFron			TIVOre			TIVCos			TIVFlan		
			Max	Med	Min	Max	Med	Min	Max	Med	Min	Max	Med	Min	Max	Med	Min
FR	1,00	0,42	0,48	0,36	0,38	0,33	0,38	0,46	0,50	0,45	0,46	0,41	0,39	0,41	0,38		
TR		1,00	0,55	0,38	0,35	0,26	0,49	0,48	0,43	0,39	0,36	0,31	0,39	0,36	0,29		
TIVOI			1,00	0,66	0,64	0,54	0,64	0,64	0,57	0,61	0,59	0,53	0,59	0,59	0,51		
				1,00	0,92	0,83	0,57	0,61	0,59	0,74	0,74	0,68	0,76	0,76	0,66		
TIVFron					1,00	0,94	0,53	0,58	0,60	0,82	0,84	0,77	0,83	0,85	0,75		
						1,00	0,43	0,48	0,50	0,78	0,80	0,74	0,79	0,81	0,73		
							1,00	0,95	0,76	0,47	0,45	0,39	0,46	0,44	0,37		
TIVOre								1,00	0,90	0,54	0,53	0,47	0,51	0,51	0,44		
									1,00	0,56	0,57	0,51	0,52	0,54	0,48		
										1,00	0,97	0,88	0,93	0,93	0,81		
TIVCos											1,00	0,94	0,92	0,94	0,84		
												1,00	0,83	0,86	0,80		
													1,00	0,96	0,84		
TIVFlan														1,00	0,92		
															1,00		

Fonte: própria autoria.

Legenda: FR – frequência respiratória; TR – temperatura retal; TIVFron – temperatura de superfície da frente dos animais obtidas com termografia; TIVOI – temperatura de superfície da região ocular dos animais obtidas com termografia; TIVOre – temperatura de superfície da orelha dos animais obtidas com termografia; TIVFlan – temperatura de superfície do flanco dos animais obtidas com termografia; TIVCos – temperatura de superfície da costela dos animais obtidas com termografia. Parâmetros: Méd – média; Min – mínima; Max – máxima.

Além da análise de correlação das variáveis fisiológicas para os dados gerais, foi feita também a separação dos dados para análise de correlação dos dados da 1ª onda e da 2ª onda, conforme Tabelas 12 e 13, respectivamente.

Tabela 12 – Coeficiente de Correlação de Pearson relacionando variáveis fisiológicas dos dados da 1ª onda de calor

Variável	FR	TR	TIVOI	TIVFron			TIVOre			TIVCos			TIVFlan		
			Max	Max	Med	Min	Max	Med	Min	Max	Med	Min	Max	Med	Min
FR	1,00	0,29	0,37	0,27	0,31	0,23	0,28	0,39	0,46	0,32	0,35	0,34	0,23	0,29	0,30
TR		1,00	0,46	0,31	0,28	0,18	0,43	0,43	0,39	0,29	0,27	0,23	0,30	0,29	0,24
TIVOI	Max		1,00	0,66	0,64	0,53	0,62	0,64	0,58	0,58	0,57	0,51	0,56	0,56	0,48
	Max			1,00	0,92	0,82	0,56	0,60	0,59	0,73	0,74	0,67	0,76	0,75	0,64
TIVFron	Med				1,00	0,94	0,52	0,58	0,60	0,82	0,84	0,76	0,83	0,84	0,73
	Min					1,00	0,41	0,46	0,49	0,77	0,78	0,71	0,78	0,80	0,71
	Max						1,00	0,96	0,78	0,43	0,42	0,37	0,41	0,41	0,34
TIVOre	Med							1,00	0,91	0,50	0,51	0,46	0,47	0,49	0,43
	Min								1,00	0,54	0,56	0,51	0,50	0,53	0,47
	Max									1,00	0,98	0,88	0,92	0,94	0,83
TIVCos	Med										1,00	0,94	0,91	0,94	0,85
	Min											1,00	0,82	0,86	0,80
	Max												1,00	0,96	0,84
TIVFlan	Med													1,00	0,92
	Min														1,00

Fonte: própria autoria

Legenda: FR – frequência respiratória; TR – temperatura retal; TIVFron – temperatura de superfície da frente dos animais obtidas com termografia; TIVOI – temperatura de superfície da região ocular dos animais obtidas com termografia; TIVOre – temperatura de superfície da orelha dos animais obtidas com termografia; TIVFlan – temperatura de superfície do flanco dos animais obtidas com termografia; TIVCos – temperatura de superfície da costela dos animais obtidas com termografia. Parâmetros: Méd – média; Min – mínima; Max – máxima.

Tabela 13 – Coeficiente de Correlação de Pearson relacionando variáveis fisiológicas dos dados da 2ª onda de calor

Variável	FR	TR	TIVOI	TIVFron			TIVore			TIVCos			TIVFlan		
			Max	Max	Med	Min	Max	Med	Min	Max	Med	Min	Max	Med	Min
FR	1,00	0,49	0,54	0,42	0,44	0,39	0,45	0,54	0,58	0,49	0,50	0,45	0,46	0,47	0,43
TR		1,00	0,63	0,43	0,40	0,33	0,55	0,54	0,47	0,45	0,42	0,36	0,45	0,41	0,33
TIVOI	Max		1,00	0,65	0,63	0,55	0,65	0,65	0,57	0,62	0,60	0,54	0,61	0,60	0,53
	Max			1,00	0,91	0,83	0,58	0,62	0,59	0,73	0,74	0,68	0,76	0,76	0,67
TIVFron	Med				1,00	0,95	0,53	0,58	0,59	0,83	0,85	0,79	0,83	0,85	0,76
	Min					1,00	0,45	0,49	0,51	0,80	0,83	0,76	0,81	0,82	0,74
	Max						1,00	0,95	0,74	0,50	0,48	0,41	0,50	0,47	0,39
TIVore	Med							1,00	0,89	0,57	0,55	0,48	0,55	0,54	0,46
	Min								1,00	0,58	0,58	0,51	0,55	0,55	0,48
	Max									1,00	0,97	0,87	0,93	0,92	0,79
TIVCos	Med										1,00	0,94	0,92	0,93	0,83
	Min											1,00	0,83	0,86	0,79
	Max												1,00	0,96	0,84
TIVFlan	Med													1,00	0,93
	Min														1,00

Fonte: própria autoria

Legenda: FR – frequência respiratória; TR – temperatura retal; TIVFron – temperatura de superfície da frente dos animais obtidas com termografia; TIVOI – temperatura de superfície da região ocular dos animais obtidas com termografia; TIVore – temperatura de superfície da orelha dos animais obtidas com termografia; TIVFlan – temperatura de superfície do flanco dos animais obtidas com termografia; TIVCos – temperatura de superfície da costela dos animais obtidas com termografia. Parâmetros: Méd – média; Min – mínima; Max – máxima.

Como pode-se observar nas Tabelas 12 e 13, a correlação da FR e TR com os valores das TIVs foi maior para os dados da 2ª onda do que para os da 1ª onda, o que favorece a construção de modelos classificadores mais precisos. Além disso, observa-se na Tabela 11 que as duas regiões que obtiveram maiores correlações com FR e TR foram as regiões do olho e da orelha. Portanto, definiu-se quatro subconjuntos de dados para treinamento dos modelos, denotados por A, B, C e D, contendo cada um deles como dados de entrada as variáveis meteorológicas (temperatura do ar e umidade relativa) e mais:

- A: Base de dados completa da região do olho;
- B: Base de dados completa da região da orelha;
- C: Base de dados do olho com dados da 1ª onda excluídos;
- D: Base de dados da orelha com dados da 1ª onda excluídos.

Os quatro subconjuntos de dados compõem um dos três parâmetros primários a serem variados na construção dos modelos para posterior comparação entre si.

### 6.3 MODELAGEM E COMPARAÇÃO DOS MODELOS

Os três algoritmos identificados pelo bloco Auto Classifier como tendo melhor desempenho foram RNA, Árvore de Decisão e Linear SVM. O número de modelos construídos foi resultante da combinação dos três parâmetros primários:

- Subconjunto de dados: 4 subconjuntos (A, B, C e D);
- Tipo de rotulagem de dados: 3 tipos (C2/FR, C3/FR e C2/TR);
- Algoritmo de classificação: 3 algoritmos (RNA, Árvore de Decisão e Linear SVM).

Portanto, foram construídos e avaliados 36 modelos a partir das diferentes combinações descritas anteriormente. As acurácias obtidas pelos modelos estão apresentadas na Tabela 14.

Tabela 14 – Acurácia dos modelos

				Continua
Modelo	Subconjunto de dados	Tipo de rotulagem	Algoritmo de classificação	Acurácia
1	A	2C/TR	RNA	71,7%
2	A	2C/TR	Árvore de decisão	68,3%
3	A	2C/TR	Linear SVM	68,0%
4	A	2C/FR	RNA	85,1%
5	A	2C/FR	Árvore de decisão	78,7%
6	A	2C/FR	Linear SVM	85,0%
7	A	3C/FR	RNA	53,9%
8	A	3C/FR	Árvore de decisão	54,0%
9	A	3C/FR	Linear SVM	53,5%
10	B	2C/TR	RNA	71,7%
11	B	2C/TR	Árvore de decisão	68,0%
12	B	2C/TR	Linear SVM	69,2%
13	B	2C/FR	RNA	86,8%
14	B	2C/FR	Árvore de decisão	81,3%

Modelo	Subconjunto de dados	Tipo de rotulagem	Algoritmo de classificação	Conclusão
				Acurácia
15	B	2C/FR	Linear SVM	84,9%
16	B	3C/FR	RNA	56,7%
17	B	3C/FR	Árvore de decisão	55,0%
18	B	3C/FR	Linear SVM	55,7%
19	C	2C/TR	RNA	76,0%
20	C	2C/TR	Árvore de decisão	73,4%
21	C	2C/TR	Linear SVM	73,9%
22	C	2C/FR	RNA	85,5%
23	C	2C/FR	Árvore de decisão	82,1%
24	C	2C/FR	Linear SVM	84,3%
25	C	3C/FR	RNA	62,6%
26	C	3C/FR	Árvore de decisão	61,9%
27	C	3C/FR	Linear SVM	60,6%
28	D	2C/TR	RNA	75,3%
29	D	2C/TR	Árvore de decisão	72,9%
30	D	2C/TR	Linear SVM	74,3%
31	D	2C/FR	RNA	85,8%
32	D	2C/FR	Árvore de decisão	81,8%
33	D	2C/FR	Linear SVM	85,7%
34	D	3C/FR	RNA	64,4%
35	D	3C/FR	Árvore de decisão	60,1%
36	D	3C/FR	Linear SVM	63,4%

Fonte: própria autoria

Legenda: Subconjunto de dados: A – base de dados completa da região do olho, B – base de dados completa da região da orelha, C – base de dados do olho com dados da 1ª onda excluídos e D – base de dados da orelha com dados da 1ª onda excluídos. Tipo de rotulagem: 2C/TR – rotulagem baseada em TR e feita em 2 classes, 2C/FR – rotulagem baseada em FR e feita em 2 classes, 3C/FR – rotulagem baseada em FR e feita em 3 classes.

As acurácias normalizadas obtidas pelos modelos estão apresentadas na Tabela 15. Foi utilizada uma escala de cores para facilitar a visualização, na qual os níveis mais escuros de verde representam maiores valores de acurácia normalizada.

Tabela 15 – Acurácia normalizada dos modelos

Subconjunto de dados	Algoritmo de classificação e tipo de rotulagem								
	RNA			Árvore de Decisão			Linear SVM		
	2C/TR	2C/FR	3C/FR	2C/TR	2C/FR	3C/FR	2C/TR	2C/FR	3C/FR
A	43,4%	70,2%	30,9%	36,6%	57,4%	31,0%	36,0%	70,0%	30,3%
B	43,4%	73,6%	35,1%	36,0%	62,6%	32,5%	38,4%	69,8%	33,6%
C	52,0%	71,0%	43,9%	46,8%	64,2%	42,9%	47,8%	68,6%	40,9%
D	50,6%	71,6%	46,6%	45,8%	63,6%	40,2%	48,6%	71,4%	45,1%

Fonte: própria autoria.

Legenda: Subconjunto de dados: A – base de dados completa da região do olho, B – base de dados completa da região da orelha, C – base de dados do olho com dados da 1ª onda excluídos e D – base de dados da orelha com dados da 1ª onda excluídos. Tipo de rotulagem: 2C/TR – rotulagem baseada em TR e feita em 2 classes, 2C/FR – rotulagem baseada em FR e feita em 2 classes, 3C/FR – rotulagem baseada em FR e feita em 3 classes. Escala de cores – níveis mais escuros de verde representam maiores valores de acurácia.

Analisando-se as acurácias dos modelos, pode-se notar que de fato os subconjuntos de dados com dados de 1ª onda excluídos obtiveram melhores acurácias que os que não tiveram os dados de 1ª onda excluídos. Por outro lado, o modelo que obteve maior acurácia (modelo 13) foi construído com dados completos e é uma exceção.

Nota-se também que modelos com rotulagem do tipo 2C/FR obtiveram melhores resultados que modelos de 2C/TR e de 3C/FR. Os modelos que classificam em 2 classes obtiveram maiores acurácias normalizadas que os que classificam em 3 classes, o que mostra menor eficácia da classificação para o segundo caso em comparação com o primeiro.

Comparando-se as regiões do olho e da orelha, observa-se que em tendência geral não houve diferença considerável entre as acurácias obtidas. Os 3 melhores modelos, entretanto, foram construídos a partir de dados da região da orelha.

Ao se comparar os diferentes algoritmos, percebe-se que o algoritmo de RNA obteve melhores resultados e Árvore de Decisão obteve piores resultados.

#### 6.4 COMPARAÇÃO DOS MODELOS COM O MÉTODO TRADICIONAL

Os modelos de maior acurácia que classificam o estresse térmico em 2 e 3 classes foram, respectivamente, o modelo 13 e o modelo 34. O primeiro foi criado a partir do subconjunto de dados para treinamento composto pela base de dados completa da região da orelha, sendo feita a rotulagem do tipo 2C/FR e utilizado o algoritmo RNA, enquanto que o segundo foi criado a partir do subconjunto de dados para treinamento composto pela base de dados da região da orelha com dados da 1ª onda excluídos, sendo feita a rotulagem do tipo 3C/FR e utilizado o algoritmo RNA. Nota-se que ambos utilizaram o algoritmo RNA e a rotulagem baseada em FR, bem como dados da região da orelha como subconjuntos de dados de treinamento.

A Tabela 16 mostra os hiperparâmetros utilizados na construção dos modelos.

Tabela 16 – Hiperparâmetros dos modelos de maior acurácia

<b>Modelo</b>	<b>Nº de camadas ocultas</b>	<b>Nº neurônios na camada oculta</b>	<b>Função de ativação da camada oculta</b>	<b>Função de ativação da camada de saída</b>	<b>Nº máximo de ciclos de treinamento</b>
13	1	8	Tangente hiperbólica	Softmax	250
34	1	7	Tangente hiperbólica	Softmax	250

Fonte: própria autoria.

As Tabelas 17 e 18 mostram, respectivamente, a matriz de confusão obtida para o modelo 13 e a matriz de confusão para o método tradicional de classificação em 2 classes de estresse baseado no ITU.



Tabela 17 – Matriz de confusão do Modelo 13

Observado	Predito		Sensibilidade
	Não-estresse	Estresse	
Não-estresse	162	273	37,2%
Estresse	88	2207	96,2%
Precisão	64,8%	89,0%	86,8%

Fonte: própria autoria.

Tabela 18 – Matriz de confusão gerada pelo método tradicional, 2 classes

Observado	Predito		Sensibilidade
	Não-estresse	Estresse	
Não-estresse	264	171	60,7%
Estresse	666	1629	71,0%
Precisão	28,4%	90,5%	69,3%

Fonte: própria autoria.

Pode-se observar pelas matrizes de confusão que o modelo de aprendizado de máquina obteve acurácia superior à obtida pelo método tradicional, o que mostra a utilidade do processo proposto neste trabalho. Observa-se também que ambos os modelos obtiveram maiores precisão e sensibilidade para a classe de estresse, o que se deve ao desbalanceamento da base de dados e mostra maior propensão dos modelos em classificar corretamente animais que estejam em condição de estresse. Por fim, observa-se que o modelo de aprendizado de máquina obteve maiores métricas de precisão e sensibilidade que o baseado em ITU para todas as classes.

As Tabelas 19 e 20 mostram, respectivamente, a matriz de confusão obtida para o modelo 34 e a matriz de confusão para o método tradicional de classificação em 3 classes de estresse baseado no ITU.

Tabela 19 – Matriz de confusão do Modelo 34

Observado	Predito			Sensibilidade
	Conforto	Alerta	Perigo	
Conforto	150	46	80	54,3%
Alerta	89	95	311	19,2%
Perigo	21	73	875	90,3%
<b>Precisão</b>	57,7%	44,4%	69,1%	64,4%

Fonte: própria autoria.

Tabela 20 – Matriz de confusão gerada pelo método tradicional, 3 classes

Observado	Predito			Sensibilidade
	Conforto	Alerta	Perigo	
Conforto	159	111	6	57,6%
Alerta	213	204	78	41,2%
Perigo	228	345	396	40,9%
<b>Precisão</b>	26,5%	30,9%	82,5%	43,6%

Fonte: própria autoria.

Para os modelos classificadores de 3 classes, também observa-se, assim como nos modelos de 2 classes, maior acurácia do modelo de aprendizado de máquina que do método tradicional, sendo a diferença maior aqui do que no caso anterior. A classe de Perigo foi a que obteve métricas de precisão e sensibilidade maiores no modelo de aprendizado de máquina.

## 6.5 DISCUSSÃO

As acurácias obtidas para os modelos estão próximas das obtidas em trabalhos anteriores. Sousa et al. (2016) desenvolveram um classificador de

estresse térmico de bovinos Nelore em 4 classes (conforto, alerta, perigo e emergência) baseado em lógica fuzzy a partir da TR, obtendo acurácia de 85,4%, próximo à acurácia do melhor modelo deste trabalho (86,8%). É preciso ponderar, entretanto, que o maior valor de acurácia obtido por Sousa et al. (2016) se deve ao grande desbalanceamento da base de dados, na qual cerca de 95,6% dos dados eram referentes a animais em condição de conforto térmico segundo a mensuração direta, o que se evidencia pelo baixo valor de precisão obtido para as classes de estresse (20,5% para a classe de Alerta e zero para Perigo e Emergência) e pelo fato da base de dados não possuir registros para a classe Emergência. Sousa et al. (2018), semelhante ao trabalho anterior, desenvolveram um classificador de estresse em 4 classes a partir da TR para bovinos Nelore, baseado desta vez em RNA, obtendo acurácia de 94,35%, superior à deste trabalho. Porém, os mesmos problemas do trabalho anterior foram verificados, o desbalanceamento da base de dados (94,2% dos dados referentes à classe Conforto), a ausência de animais na classe Emergência e a baixa precisão para as classes de estresse (50,7% para Alerta e zero para Perigo e Emergência).

Pacheco et al. (2020) desenvolveram dois classificadores de estresse de vacas holandesas adultas baseados em RNA, um no qual a classificação teve como referência a FR e outro teve como referência a TR, obtendo acurácias de 83% e 84%, respectivamente. O balanceamento da base de dados obtido por Pacheco et al. (2020) foi melhor que o obtido pelos trabalhos citados no parágrafo anterior, com 71,1% dos dados para a classe Conforto tendo como referência a FR e 80,0% com a TR, porém também sem dados para a classe Emergência tendo como referência a FR. Deve-se ressaltar aqui que a base de dados utilizada neste trabalho também apresentou desbalanceamento, entretanto este se verificou com maior número de dados referentes às classes de estresse, o que favorece a construção de modelos com menor propensão ao erro do tipo falso negativo (ou seja, classificar como em conforto o animal em estresse), evidenciado pela alta precisão obtida para as classes de estresse nos modelos. Na identificação de animais em condição de estresse térmico, classificar como em conforto um animal que está em estresse é mais prejudicial do que o contrário, o que mostra a maior utilidade dos modelos desenvolvidos neste trabalho se comparado aos dos trabalhos citados anteriormente

Li et al. (2020) desenvolveram um modelo para predição e classificação de estresse de bovinos holandeses, tendo como variáveis de entrada T, UR, velocidade do vento, bloco de tempo e produção de leite, obtendo acurácia de 90,5%, superior à acurácia do melhor modelo deste trabalho (86,8%). No entanto, a produção de leite não constitui um atributo prático para utilização na em um sistema de monitoramento do estresse térmico, pois não possibilita o monitoramento em tempo real e também não pode ser aplicado a vacas que não estejam em lactação. Além disso, a redução na produção de leite é um sinal atrasado em relação aos efeitos do estresse, pois indica uma condição de estresse já instalada no animal. Sistemas de monitoramento do estresse devem preferencialmente buscar monitorar a resposta imediata do animal ao ambiente estressor, o que pode ser feito pela avaliação da FR.

Becker et al. (2021) desenvolveram modelos para classificação de estresse de vacas leiteiras baseados em 3 diferentes algoritmos (floresta randômica, regressão logística e Naive Bayes gaussiano) e para 4 tipos de tratamento de conforto térmico, obtendo acurácias entre 81,1% e 89,3%. Apesar da acurácia obtida, é preciso observar que os modelos desenvolvidos por Becker et al. (2021) foram gerados tendo como variáveis de entrada um total de 18 variáveis, muitas delas tendo pouca relação com o estresse térmico ou sendo de difícil mensuração prática (como por exemplo número de passos por dia, teor de gordura e de proteína no leite, condição de higiene e condição corporal). Além disso, os modelos tiveram como variáveis de entrada a FR e o fato de a vaca estar babando ou não, que são variáveis utilizadas também na rotulagem dos dados (denominado pelos autores de sistema de escores de estresse térmico), o que faz os modelos terem pouca utilidade por não serem feitos para obter informação nova sobre variáveis das quais não se tem conhecimento prévio.

## 7 CONCLUSÃO

Neste trabalho foi desenvolvido e avaliado um processo de descoberta de conhecimento para classificação de estresse térmico de bezerras holandeses a partir de dados ambientais e de termografia. O processo proposto buscou construir modelos a partir do ajuste de diferentes parâmetros e comparar os resultados. Os resultados mostram que a assinatura térmica é efetiva como atributo preditor do estresse e que o método proposto é capaz de gerar modelos com bom desempenho. Foi possível confirmar a hipótese levantada inicialmente, obtendo-se modelos com alta acurácia na classificação do estresse, com resultados semelhantes ao obtido por outros trabalhos, porém utilizando atributos previsores não-invasivos e mais práticos quando se considera a aplicação em sistemas de monitoramento individual e em tempo real.

Apesar de este trabalho ter gerado bons modelos, acredita-se que o desempenho deles pode ser melhorado em trabalhos subsequentes com, por exemplo, ampliação da base de dados com maior balanceamento entre as classes de estresse e conforto térmico. Outro fator do projeto foi a ausência de mecanismos de resfriamento na câmara climática, que poderia permitir o controle refinado de temperatura para condições de estresse e conforto. Além disso, a metodologia proposta neste trabalho pode ser explorada em trabalhos com outras raças e espécies de animais.

## REFERÊNCIAS

- ABDUL JABBAR, K. et al. Early and non-intrusive lameness detection in dairy cows using 3-dimensional video. **Biosystems Engineering**, v. 153, p. 63–69, jan. 2017.
- ALSAAOD, M. et al. A field trial of infrared thermography as a non-invasive diagnostic tool for early detection of digital dermatitis in dairy cows. **The Veterinary Journal**, v. 199, n. 2, p. 281–285, fev. 2014.
- ARCIDIACONO, C.; MANCINO, M.; PORTO, S. M. C. Moving mean-based algorithm for dairy cow's oestrus detection from uniaxial-accelerometer data acquired in a free-stall barn. **Computers and Electronics in Agriculture**, v. 175, p. 105498, ago. 2020.
- ASTILL, J. et al. Smart poultry management: Smart sensors, big data, and the internet of things. **Computers and Electronics in Agriculture**, v. 170, p. 105291, mar. 2020.
- BECKER, C. A.; COLLIER, R. J.; STONE, A. E. Invited review: Physiological and behavioral effects of heat stress in dairy cows. **Journal of Dairy Science**, v. 103, n. 8, p. 6751–6770, 1 ago. 2020.
- BECKER, C. A. et al. Predicting dairy cattle heat stress using machine learning techniques. **Journal of Dairy Science**, v. 104, n. 1, p. 501–524, 1 jan. 2021.
- BENAISSA, S. et al. Calving and estrus detection in dairy cattle using a combination of indoor localization and accelerometer sensors. **Computers and Electronics in Agriculture**, v. 168, p. 105153, jan. 2020.
- BERRY, R. J. et al. Daily variation in the udder surface temperature of dairy cows measured by infrared thermography: Potential for mastitis detection. **Canadian Journal of Animal Science**, v. 83, n. 4, p. 687–693, 1 dez. 2003.
- BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. **Projeções do Agronegócio: Brasil 2019/20 a 2029/2030**. Brasília: MAPA, 2020.
- CARVAJAL, M. A. et al. Increasing importance of heat stress for cattle farming under future global climate scenarios. **Science of The Total Environment**, v. 801, p. 149661, dez. 2021.
- CHURCH, J. S. et al. Influence of environmental factors on infrared eye temperature measurements in cattle. **Research in Veterinary Science**, v. 96, n. 1, p. 220–226, fev. 2014.
- COMPANHIA NACIONAL DE ABASTECIMENTO. **Pecuária leiteira: análise dos custos de produção e da rentabilidade nos anos de 2014 a 2017**. Brasília: Conab, 2018. Disponível em <[https://www.conab.gov.br/institucional/publicacoes/compendio-de-estudos-da-](https://www.conab.gov.br/institucional/publicacoes/compendio-de-estudos-da)

conab/item/download/23676\_f8c5e386b295537c601b8671c70993fc>. Acesso em: 20 mai. 2020.

DALTRO, D. DOS S. et al. Infrared thermography as a method for evaluating the heat tolerance in dairy cows. **Revista Brasileira de Zootecnia**, v. 46, n. 5, p. 374–383, maio 2017.

DIAS BATISTA, P. H. et al. Thermal images to predict the thermal comfort index for Girolando heifers in the Brazilian semiarid region. **Livestock Science**, v. 251, p. 104667, 1 set. 2021.

DUKES, H. H. **Fisiologia dos animais domésticos**. 11. ed. Rio de Janeiro: Guanabara, Koogan, 1996. 856 p.

FAO: banco de dados. Disponível em: <<http://www.fao.org/faostat/en/#data> >. Acesso em: 14 jul. 2022.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From Data Mining to Knowledge Discovery in Databases. **AI Magazine**, v. 17, n. 3, p. 37, 15 Mar. 1996.

FERREIRA, I. C. et al. **Conforto Térmico em Bovinos Leiteiros a Pasto**. Planaltina, DF: Embrapa Cerrados, ISSN 1517-5111, ISSN online 2176-5081, Documentos 342, set. 2017. Disponível em: <<https://www.infoteca.cnptia.embrapa.br/infoteca/bitstream/doc/1092310/1/Doc342.pdf>>. Acesso em: 20 out. 2020.

GIRO, A. et al. Application of microchip and infrared thermography for monitoring body temperature of beef cattle kept on pasture. **Journal of Thermal Biology**, v. 84, p. 121–128, ago. 2019.

GUO, Y.; HE, D.; SONG, H. Region detection of lesion area of knee based on colour edge detection and bilateral projection. **Biosystems Engineering**, v. 173, p. 19–31, set. 2018.

HAHN, G.L.; MADER, T.L. Heat waves in relation on thermoregulation, feeding behavior, and mortality of feedlot cattle. In : International Livestock 42 Environment Symposium, 5., Mennesota, 1997. Proceedings. ST. Joseph: ASAE, 1997. p.125-129.

HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. **The Elements of Statistical Learning: Data Mining, Inference and Prediction**. 2nd ed. Stanford: Springer, 2017. Disponível em: <<https://web.stanford.edu/~hastie/ElemStatLearn/>>. Acesso em: 01 out. 2020.

HAYKIN, S. **Redes Neurais: Princípios e Prática**. 2. Ed. Tradução de Paulo Martins Engel. São Paulo: Editora Artmed LTDA, 1999.

HOFFMANN, G. et al. Animal-related, non-invasive indicators for determining heat stress in dairy cows. **Biosystems Engineering**, 8 nov. 2019.

IBGE: banco de dados. Disponível em: <<https://www.ibge.gov.br/estatisticas/economicas/agricultura-e-pecuaria/21121-primeiros-resultados-2leite.html?edicao=26867&t=series-historicas>>. Acesso em: 14 jul. 2022.

IBM Watson Studio – Data Science e IA. Disponível em: <<https://www.ibm.com/br-pt/cloud/watson-studio>>. Acesso em: 10 maio 2020.

JOHNSON, H. D.; VANJONACK, W. J. Effects of Environmental and Other Stressors on Blood Hormone Patterns in Lactating Animals. **Journal of Dairy Science**, v. 59, n. 9, p. 1603–1617, set. 1976.

KAMILARIS, A.; KARTAKOULLIS, A.; PRENAFETA-BOLDÚ, F. X. A review on the practice of big data analysis in agriculture. **Computers and Electronics in Agriculture**, v. 143, p. 23–37, dez. 2017.

KNÍŽKOVÁ, I. et al. Applications of infrared thermography in animal production. **J. Fac. Agric.**, v. 22, p. 329-336. 2007

KOTU, V.; DESHPANDE, B. Classification. In: **Predictive Analytics and Data Mining**. [s.l.] Elsevier, 2015. p. 63–163, 257–273.

KRIEGER, S. et al. Prediction of calving in dairy cows using a tail-mounted tri-axial accelerometer: A pilot study. **Biosystems Engineering**, v. 173, p. 79–84, set. 2018.

LI, G. et al. Predicting rectal temperature and respiration rate responses in lactating dairy cows exposed to heat stress. **Journal of Dairy Science**, v. 103, n. 6, p. 5466–5484, 1 jun. 2020.

MARTELLO, L. S. **Interação animal-ambiente**: efeito do ambiente climático sobre as respostas fisiológicas e produtivas de vacas Holandesas em free-stall. 2006. Tese (Doutorado) - Faculdade de Zootecnia e Engenharia de Alimentos, Universidade de São Paulo, Pirassununga, 2006.

MARTINS, R. F. S. et al. Mastitis detection in sheep by infrared thermography. **Research in Veterinary Science**, v. 94, n. 3, p. 722–724, jun. 2013.

MATTACHINI, G. et al. Monitoring feeding behaviour of dairy cows using accelerometers. **Journal of Agricultural Engineering**, v. 47, n. 1, p. 54, 8 mar. 2016.

MCMANUS, C. et al. Infrared thermography in animal production: An overview. **Computers and Electronics in Agriculture**, v. 123, p. 10–16, abr. 2016.

MOROTA, G. et al. BIG DATA ANALYTICS AND PRECISION ANIMAL AGRICULTURE SYMPOSIUM: Machine learning and data mining advance predictive big data analysis in precision animal agriculture1. **Journal of Animal Science**, v. 96, n. 4, p. 1540–1550, 14 abr. 2018.



NÃÃS, I.A. **Princípios de conforto térmico na produção animal**. São Paulo: Ícone, 1989.

NAVON, S. et al. Automatic recognition of jaw movements in free-ranging cattle, goats and sheep, using acoustic monitoring. **Biosystems Engineering**, v. 114, n. 4, p. 474–483, abr. 2013.

NEETHIRAJAN, S. The role of sensors, big data and machine learning in modern animal farming. **Sensing and Bio-Sensing Research**, v. 29, p. 100367, ago. 2020.

NIKKHAH, A. et al. Short Communication: Infrared Thermography and Visual Examination of Hooves of Dairy Cows in Two Stages of Lactation. **Journal of Dairy Science**, v. 88, n. 8, p. 2749–2753, ago. 2005.

NIR, O. et al. 3D Computer-vision system for automatically estimating heifer height and body mass. **Biosystems Engineering**, v. 173, p. 4–10, set. 2018.

OUDSHOORN, F. W. et al. Estimation of grass intake on pasture for dairy cows using tightly and loosely mounted di- and tri-axial accelerometers combined with bite count. **Computers and Electronics in Agriculture**, v. 99, p. 227–235, nov. 2013.

PACHECO, V. M. et al. Thermal imaging combined with predictive machine learning based model for the development of thermal stress level classifiers. **Livestock Science**, v. 241, p. 104244, nov. 2020.

PAHL, C. et al. Suitability of feeding and chewing time for estimation of feed intake in dairy cows. **animal**, v. 10, n. 9, p. 1507–1512, set. 2016.

POLAT, B. et al. Sensitivity and specificity of infrared thermography in detection of subclinical mastitis in dairy cows. **Journal of Dairy Science**, v. 93, n. 8, p. 3525–3532, ago. 2010.

POLSKY, L.; VON KEYSERLINGK, M. A. G. Invited review: Effects of heat stress on dairy cattle welfare. **Journal of Dairy Science**, v. 100, n. 11, p. 8645–8657, nov. 2017.

RAINWATER-LOVETT, K. et al. Detection of foot-and-mouth disease virus infected cattle using infrared thermography. **The Veterinary Journal**, v. 180, n. 3, p. 317–324, jun. 2009.

RICCI, G. D.; SILVA-MIRANDA, K. O. DA; TITTO, C. G. Infrared thermography as a non-invasive method for the evaluation of heat stress in pigs kept in pens free of cages in the maternity. **Computers and Electronics in Agriculture**, v. 157, p. 403–409, fev. 2019.

ROBINSON E.N. **Termorregulação**. In: CUNNINGHAM, J.G. Tratado de fisiologia veterinária. 2.ed. Rio de Janeiro: Guanabara Koogan, cap.51, 427-435, 1999.

RODRIGUES, A. V. S. **Avaliação não invasiva do estresse térmico de bovinos: uma abordagem baseada em aprendizado de máquina e termografia de infravermelho.** 2022. 62 f. Dissertação (Mestrado) - Faculdade de Zootecnia e Engenharia de Alimentos. Universidade de São Paulo, Pirassununga, 2022.

ROSENBERG, N. J.; BLAD, B. L.; VERMA, S. B. Human and Animal Biometeorology. In: **Microclimate: the biological environment.** 2nd ed ed. New York: Wiley, 1983.

SALLES, M. S. V. et al. Mapping the body surface temperature of cattle by infrared thermography. **Journal of Thermal Biology**, v. 62, p. 63–69, dez. 2016.

SCHAEFER, A. L. et al. Early detection and prediction of infection using infrared thermography. **Canadian Journal of Animal Science**, v. 84, n. 1, p. 73–80, 1 mar. 2004.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. **Information Processing & Management**, v. 45, n. 4, p. 427–437, jul. 2009.

SOUSA, R. V. DE et al. Development and evaluation of a fuzzy logic classifier for assessing beef cattle thermal stress using weather and physiological variables. **Computers and Electronics in Agriculture**, v. 127, p. 176–183, set. 2016.

SOUSA, R. V. DE et al. Predictive model based on artificial neural network for assessing beef cattle thermal stress using weather and physiological variables. **Computers and Electronics in Agriculture**, v. 144, p. 37–43, jan. 2018.

STELLETTA, C. et al. Thermographic Applications in Veterinary Medicine. In: PRAKASH, R. V. (Ed.). **Infrared Thermography.** [s.l.] InTech, 2012. Disponível em: <<https://www.intechopen.com/books/infrared-thermography/thermographic-applications-in-veterinary-medicine>>. Acesso em: 15 ago. 2020.

STEWART, M. et al. The use of infrared thermography and accelerometers for remote monitoring of dairy cow health and welfare. **Journal of Dairy Science**, v. 100, n. 5, p. 3893–3901, maio 2017.

THOM, E. C. The Discomfort Index. **Weatherwise**, v. 12, n. 2, p. 57–61, 1 abr. 1959.

UDDIN, J. et al. Relationships between body temperatures and behaviours in lactating dairy cows. **Applied Animal Behaviour Science**, v. 241, p. 105359, 1 ago. 2021.

VAN HERTEM, T. et al. Automatic lameness detection based on consecutive 3D-video recordings. **Biosystems Engineering**, v. 119, p. 108–116, mar. 2014.

WALCZAK, S.; CERPA, N. Artificial Neural Networks. In: **Encyclopedia of Physical Science and Technology**. [s.l.] Elsevier, 2003. p. 631–645.

WATHES, C. M. et al. Is precision livestock farming an engineer's daydream or nightmare, an animal's friend or foe, and a farmer's panacea or pitfall? **Computers and Electronics in Agriculture**, Smart Sensors in precision livestock farming. v. 64, n. 1, p. 2–10, 1 nov. 2008.

WILLIAMS, L. R. et al. A sensor-based solution to monitor grazing cattle drinking behaviour and water intake. **Computers and Electronics in Agriculture**, v. 168, p. 105141, jan. 2020.

WINCKLER, C. Assessing animal welfare at the farm level: Do we care sufficiently about the individual? **Animal Welfare**, v. 28, p. 77-82. 2019. 10.7120/09627286.28.1.077.

XAVIER, A. C.; KING, C. W.; SCANLON, B. R. Daily gridded meteorological variables in Brazil (1980-2013): DAILY GRIDDED METEOROLOGICAL VARIABLES IN BRAZIL (1980-2013). **International Journal of Climatology**, v. 36, n. 6, p. 2644–2659, maio 2016.