Deep Learning and Acoustic Parameter Analysis for Identifying Cattle Vocalizations under Confinement and Handling Conditions
Bryan Teixeira Paiva, Ana Paula Lüdtke Ferreira, and Naylor Bastiani Perez
SADIO Electronic Journal of Informatics and Operations Research (EJS) Vol. 24 No. 2 (2025) e-ISSN 1514-6774
https://doi.org/10.24215/15146774e077 | https://revistas.unlp.edu.ar/ejs
Sociedad Argentina de Informática e Investigación Operativa | Universidad Nacional de La Plata | Buenos Aires | Argentina

Deep Learning and Acoustic Parameter Analysis for Identifying Cattle Vocalizations under Confinement and Handling Conditions

Bryan Teixeira Paiva[1][0000−0003−0031−9970], Ana Paula Lüdtke Ferreira[1][0000−0001−7057−9095], and Naylor Bastiani Perez[1,2][0000−0002−4667−783X]

[1] Programa de Pós-graduação em Computação Aplicada
Universidade Federal do Pampa, Bagé-RS, Brazil
{bryanpaiva.aluno,anaferreira}@unipampa.edu.br
[2] Empresa Brasileira de Pesquisa Agropecuária, Bagé-RS, Brazil
naylor.perez@embrapa.br

Abstract Livestock production systems play a fundamental role in the Brazilian economy, serving as one of the country's primary sources of income. During various stages of production – including veterinary procedures, weighing, and transportation – animals are subjected to different levels of handling, with the potential to cause stress. Animal stress significantly impacts meat properties, including reducing its quality to the point of making it unfit for human consumption. This paper analyzed cattle vocalizations under two psychologically distinct conditions: confinement and handling. The main objective is to identify stress-inducing situations using two approaches: analyzing the vocalizations' acoustic parameters and applying them to train a deep learning network to learn the sound patterns emitted by the animals in each situation. In the acoustic analysis, a statistical study was conducted on the parameters of fundamental frequency (F0), spectral formants (F1–F4), jitter, shimmer, harmony, and intensity. For the deep learning study, three convolutional neural network architectures were implemented, using Mel Frequency Cepstral Coefficients (MFCC) for acoustic feature extraction. The results of the acoustic analysis revealed significant differences ($p<0.001$) between the parameters of stressed and non-stressed vocalizations for most parameters. Meanwhile, the neural network results show that the basic, intermediate, and robust architectures achieved F1-scores of 96.97%, 97.90%, and 98.74%, respectively.

Keywords: Acoustic Analysis · Deep Learning · Precision livestock

Resumen Los sistemas de producción ganadera desempeñan un papel fundamental en la economía brasileña, siendo una de las principales fuentes de ingresos del país. Durante diversas etapas de la producción, como los procedimientos veterinarios, el pesaje y el transporte, los animales son sometidos a diferentes niveles de manejo, cada uno con el potencial de causarles estrés. El estrés animal impacta significativamente las propiedades de la carne, reduciendo su calidad al punto de volverla inadecuada

Received May 2025; Accepted June 2025; Published July 2025

para el consumo humano. Este estudio analizó las vocalizaciones de bovinos en dos condiciones psicológicamente distintas: confinamiento y manejo. El objetivo principal fue identificar situaciones de estrés mediante dos enfoques: el análisis de parámetros acústicos y el uso de aprendizaje profundo aplicado a los patrones sonoros emitidos por los animales en cada situación. En el análisis acústico, se realizó un estudio estadístico de los parámetros de frecuencia fundamental (F0), formantes espectrales (F1–F4), jitter, shimmer, armonía e intensidad. En el estudio de aprendizaje profundo, se implementaron tres arquitecturas de redes neuronales convolucionales (CNN), utilizando Coeficientes Cepstrales en Frecuencia Mel (MFCC) para la extracción de características acústicas. Los resultados del análisis acústico revelaron diferencias significativas ($p<0,001$) entre los parámetros de vocalizaciones estresadas y no estresadas en la mayoría de los casos. Por su parte, los resultados de las redes neuronales mostraron que las arquitecturas básica, intermedia y robusta lograron puntajes F1 del 96,97%, 97,90% y 98,74%, respectivamente.

Palabras clave: Análisis acústico · Aprendizaje profundo · Ganadería de precisión

1 Introdução

A pecuária bovina de corte é uma das principais atividades geradoras de renda no Brasil, com produção estimada em 213 milhões de cabeças de gado [13]. O setor é responsável por 6% do PIB brasileiro, sendo um dos principais produtores, consumidores e exportadores de carne do mundo.

Ainda que possua uma produção expressiva, o setor produtivo da pecuária de corte está continuamente em busca de formas de aumentar sua lucratividade e melhorar a qualidade da carne produzida, tanto para atender aos padrões rigorosos exigidos para a exportação para mercados de alto valor agregado [5] quanto para atender a um perfil de consumidor mais preocupado com questões que envolvem a criação e o abate de animais.

Nos últimos anos, nota-se uma mudança no perfil de consumidores comuns, com uma maior exigência em relação à qualidade da carne que consomem. Além dos atributos tradicionais como aparência, sabor, textura e aroma, há um número crescente de consumidores preocupados com o bem-estar dos animais e com as práticas de criação. Nos mercados como o da União Europeia e entre consumidores com maior poder aquisitivo, há uma disposição para pagar mais por produtos que garantam um tratamento adequado aos animais ao longo de todo o processo de produção [24].

O bem-estar dos animais envolve diversos aspectos, como liberdade, conforto, estresse, medo e saúde. No contexto da criação de gado, o manejo dos animais durante as diferentes etapas de produção afeta negativamente seu bem-estar e pode prejudicar a qualidade final da carne. Um manejo inadequado pode causar desde lesões leves, como hematomas, até situações extremas que levam ao óbito dos animais, acarretando perdas financeiras para os produtores e frigoríficos, além do sofrimento dos próprios animais [5].

O estresse sofrido pelos animais durante o manejo pode levar ao aumento do pH da carne, resultando em características conhecidas como DFD (escura, dura e seca, do Inglês *dark, firm, dry*). Isso compromete a qualidade da carne e gera prejuízos financeiros para os diferentes agentes da cadeia de produção - produtores, transportadores e frigoríficos - pois uma carne com essas características não é adequada para consumo humano [16].

Como exemplo de perdas financeiras decorrentes de estresse e lesões, na Austrália a incidência de carne escura em carcaça de bovinos é de 10%, causando prejuízos financeiros estimados em aproximadamente AU$ 36 milhões por ano para a indústria bovina (R$ 134,18 milhões, em valores de janeiro de 2025) [12]. No Brasil, estudos verificaram a ocorrência de lesões em 42,4% das carcaças de bovinos devido ao transporte e manejo, acarretando em perdas financeiras que podem ultrapassar R$ 200 mil por ano para um frigorífico de porte médio [28]. Esse cenário destaca a urgência de reformular os métodos empregados no manejo animal e de desenvolver tecnologias avançadas capazes de identificar automaticamente o estresse. Essas inovações não apenas promovem melhores condições de bem-estar para os animais, mas também contribuem para a maior lucratividade dos produtores.

Em estudos sobre estresse animal, as vocalizações do gado bovino têm sido examinadas como possíveis indicadores de bem-estar, usualmente por meio da análise da estrutura acústica e das informações nelas codificadas. [35]. Estudos recentes têm demonstrado que as vocalizações podem conter uma riqueza de informações sobre o estado emocional, saúde e comportamento dos animais.

Este trabalho tem como objetivo investigar a capacidade de uma rede neural em identificar situações de estresse com base nas vocalizações do gado bovino. O trabalho foi conduzido com base nas seguintes hipóteses de pesquisa: (i) os sons emitidos por animais estressados são diferentes daqueles emitidos por animais mais tranquilos; (ii) a análise dos sons emitidos pelos animais pode ser avaliada por meio de redes neurais e análise estatística de parâmetros acústicos para identificação de situações de estresse.

Para comprovação das hipóteses de pesquisa, foi conduzida uma análise estatística dos principais parâmetros acústicos avaliados na literatura concernente ao estudo e identificação de vocalizações e estresse de animais. Os parâmetros de análise acústica foram a frequência fundamental (F0), os quatro primeiros formantes do espectro das vocalizações (F1, F2, F3 e F4), *jitter*, *shimmer*, harmonia e intensidade. O trabalho foi conduzido a partir de análise de variância por testes t de Student entre os parâmetros selecionados, com dados de vocalização provenientes de dois grupos independentes (confinamento e manejo), com análise das imagens colhidas para confirmar a existência/inexistência de situação de estresse na vocalização coletada.

O restante deste trabalho está organizado como se segue: a Seção 3 apresenta o conjunto de materiais e métodos utilizados neste trabalho, a Seção 4 apresenta e discute os resultados obtidos e a Seção 5 apresenta a conclusão do trabalho, com indicação de trabalhos futuros.

2 Trabalhos correlatos

O desenvolvimento desta pesquisa e a definição dos métodos e técnicas empregados foram baseados em uma revisão de escopo da literatura [1], com vistas a entender o estado da arte e compará-lo à proposta apresentada. O protocolo de revisão [26] selecionou trabalhos relacionados às abordagens de a análise estatística e a aplicação de redes neurais (CNN) na identificação de estados comportamentais dos animais a partir de suas vocalizações.

Considerando somente análises de áudio, a literatura apresenta trabalhos que exploram a viabilidade de avaliar o bem-estar de bovinos por meio da análise dos sons emitidos, identificando comportamentos relacionados à ruminação, alimentação, interação social, interação sexual, estresse e outras atividades [23,15,7].

Técnicas de análise acústica foram usadas para identificação do cio em bovinos [3,29,19,38] e para monitoramento do comportamento ingestivo de bovinos, avaliando aspectos como mastigação e deglutição [2,4].

No contexto mais amplo da análise acústica de vocalizações, pesquisas como as de [35,37] examinaram a estrutura e as características acústicas das vocalizações bovinas por meio de análises estatísticas nos domínios da amplitude e frequência dos sinais. No que se refere à avaliação do estresse animal, estudos identificaram alterações nas vocalizações de porcos e aves como indicadores de condições estressantes [25,22,18]. Com relação aos bovinos, trabalhos como os de [14,8] analisaram mudanças nas vocalizações de vacas em situações de estresse psicológico.

Quanto à aplicação de redes neurais na análise de vocalizações de animais, diversos estudos fazem uso de CNN para automatizar a identificação de vocalizações [36,30,27,31,17]. No que diz respeito às características acústicas utilizadas como entrada para redes neurais, destaca-se o uso de coeficientes cepstrais de Mel (MFCC) em estudos como os de [30,17,31]. Por outro lado, trabalhos como [36,27,32] utilizaram espectrogramas como insumo principal para CNN, evidenciando a eficácia dessa abordagem na modelagem de vocalizações.

A Fig. 1 apresenta uma síntese dos trabalhos correlatos, onde se ressalta o foco de aplicação, a proveniência dos dados utilizados, o tipo de análise realizada e os principais resultados obtidos dos trabalhos. Dentre as principais características dos trabalhos é possível destacar o foco, o qual está compreendido em identificação de cio, análise de bem-estar, análise de comportamento e classificação de vocalizações. Quanto às técnicas utilizadas no desenvolvimento dos trabalhos, é possível destacar a utilização do algoritmo MFCC para a análise e processamento de áudio, algoritmos de aprendizado de máquina para a classificação dos sons e análises de variância para comparação das propriedades e características acústicas das vocalizações.

A comparação entre este trabalho e os trabalhos correlatos destaca algumas similaridades: a utilização de técnicas de processamento de áudio, uso de redes neurais, análise estatística nos domínios da frequência e amplitude, classificação e identificação de estresse. Contudo, este trabalho se destaca principalmente no objetivo de realizar um estudo comparativo entre diferentes arquiteturas de redes neurais para um modelo de classificação capaz de identificar situações de

Aplicação	Proveniência dos dados	Tipo de análise	Principais resultados	Trabalhos
Análise de bem-estar	Coleta de vocalizações em condições normais e estressantes	Análise de variância (testes de comparações múltiplas de Tukey)	Menor pico de frequência em condições de estresse	(MOURA et al., 2008)
	Coleta de vocalizações em condições normais e estressantes	Análise de som (LPC e redes neurais)	Maior incidência de estresse em regime de disputa por alimento	(MANTEUFFEL; SCHÖN, 2002)
	Coleta de vocalizações de condições normais e estressantes	Análise de som (CFS e SVM)	Taxa de detecção de estresse de 86,6%	(LEE et al., 2015)
	Coleta de vocalizações em condições estressantes	Análise de som (LPC)	Maiores frequências médias em condições de estresse	(IKEDA; ISHII, 2008)
Análise de comportamento	Coleta de vocalizações de animais	Análise de som (MFCC e HMM)	Taxa de identificação de estado entre 74-100%	(JAHNS, 2008)
	Coleta de vocalizações em diferentes estados	Análise de som (MFCC)	Taxa de identificação de estado entre 71-92%	(DESHMUKH et al., 2012)
	Coleta de áudio e vídeo de animais em campo	Análise de variância (teste de Fischer)	Maior frequência máxima média em situações de estresse	(MEEN et al., 2015)
	Coleta de vocalizações de animais em confinamento	Análise de som (MFCC) e redes neurais (CNN)	Taxa de classificação de estado entre 74-95%	(JUNG et al., 2021)
Classificação de vocalizações	Coleta de vocalizações em ambiente livre	Análise de variância (ANOVA, ANCOVA)	Diferenças significativas entre vocalizações de vacas e bezerros	(TORRE et al., 2015)
	Coleta de vocalizações de animais em confinamento	Redes neurais (RNN e CNN) e análise estatística	Taxa de classificação de vocalizações entre 68-87%	(GAVOJDIAN et al., 2023)
	Vocalizações reais e sintéticas	Análise de som (espectro Mel) e redes neurais (FRCNN)	Taxa de classificação de vocalizações entre 65-89%	(PANDEYA et al., 2022)

Figura 1. Análise comparativa de trabalhos correlatos

estresse, além de conduzir uma análise estatística sobre parâmetros nos domínios da frequência e amplitude para avaliar os principais fatores relacionados ao estado de estresse animal.

3 Material e métodos

O trabalho foi realizado a partir de uma metodologia dividida em seis etapas procedimentais: (i) revisão de escopo da literatura sobre o estado da arte na análise acústica e identificação de vocalizações de animais de corte (com ênfase em gado bovino); (ii) coleta de dados de vocalizações em condições de confinamento e manejo; (iii) tratamento dos dados e preparação da base de vocalizações; (iv) desenvolvimento da análise acústica; (v) implementação e treinamento das redes neurais; (vi) análise dos resultados.

3.1 Coleta de dados

A captação das vocalizações dos animais foi conduzida em dois contextos distintos: confinamento e manejo. Cada contexto representa um momento em que

diferentes níveis de estresse animal podem ser observados. As coletas foram realizadas com 48 animais da raça Brangus (*Bos taurus indicus*).

No ambiente de confinamento, situado na zona rural da cidade de Bagé/RS, Brasil (31°18'56"S 53°59'54"W), os animais estavam divididos em dois espaços de aproximadamente 1500m^2 no total (30m x 25m cada potreiro), onde podiam conviver, interagir e se alimentar livremente. Neste período, as interações com humanos eram minimizadas, ocorrendo somente quando era necessário repor a alimentação dos animais, garantindo assim a manutenção de um ambiente controlado e livre de estresse. Para a captação dos sons foram instaladas três câmeras para cobertura total do ambiente: uma na extremidade esquerda, outra central e a terceira na extremidade direita do local. As câmeras foram posicionadas próximas aos cochos, locais nos quais os animais se concentravam na maior parte de seu tempo. A Fig. 2 apresenta a disposição das câmeras para coleta de vocalizações de confinamento.

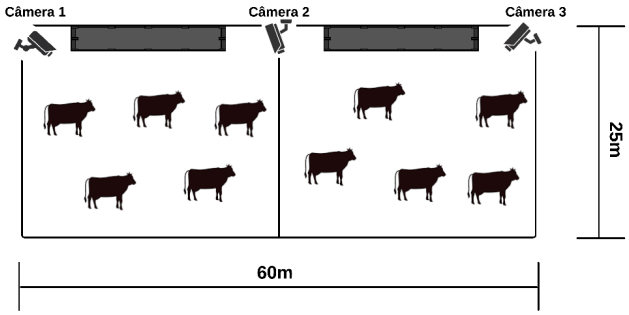


Figura 2. Disposição das câmeras no ambiente de confinamento

As imagens foram registradas ao longo de 14 dias, totalizando aproximadamente 1000 horas de filmagens. As vocalizações registradas nesse ambiente foram categorizadas como livres de estresse, refletindo o estado sereno dos animais e a ausência de interações estressantes.

A coleta de vocalizações dos animais em situação de manejo ocorreu durante do processo de pesagem dos animais. Nesse cenário observou-se um manejo caracterizado por interações enérgicas que envolviam gritos, gestos e cutucões. Foram acompanhadas duas sessões de pesagem, com duração aproximada de uma hora cada. Durante o manejo, os animais foram conduzidos para um curral estreito que os levava até a balança. Nesse percurso, a interação entre humanos e animais tornava-se mais explícita, culminando na contenção dos animais dentro da gaiola de pesagem até a realização da leitura do peso. Após esse procedimento, o animal era libertado da gaiola, abrindo espaço para a entrada do próximo animal a ser pesado. A Fig. 3 apresenta os locais de confinamento (à esquerda) e de manejo dos animais (à direita).

Figura 3. Animais em confinamento (esq) e em manejo (dir)

Durante os procedimentos de manejo foi evidente a agitação e um aumento nos níveis de estresse dos animais, também atestado por especialistas em comportamento animal. O processo foi acompanhado por uma maior produção de vocalizações em comparação com os períodos de confinamento. O monitoramento e registro deste processo foi executado por meio de câmeras digitais.

O uso de câmeras, tanto do manejo como no confinamento, possibilitou uma análise mais detalhada das interações e comportamentos dos animais. A categorização de situações de estresse e não estresse foi feita por meio das imagens coletadas, em que pode-se verificar o comportamento do animal enquanto vocalizava.

3.2 Preparação da base de dados

A construção da base de dados de vocalizações e rotulação dos dados originou-se das filmagens capturadas durante a fase de coleta. Nesse estágio, foi necessária a análise dos registros e a categorização das vocalizações com base nos níveis de estresse manifestados pelos animais. Para automatizar a preparação da base de vocalizações de confinamento, foi desenvolvido um *software* em Python (https://www.python.org/) para detecção de picos de amplitude sonora nas filmagens. Como as vocalizações de bovinos geralmente variam entre 1,3 e 2,1 segundos de duração [35], o funcionamento do software envolve a criação de janelas de 3 segundos, nas quais são verificados picos de amplitude sonora. Esses picos indicam a presença de sons discrepantes e, caso identificados, os trechos correspondentes a essas janelas de tempo são individualmente armazenados para análises manuais subsequentes.

A análise manual descartou os fragmentos que não se enquadravam como vocalizações ou cujos sons não possuíam qualidade satisfatória devido à baixa amplitude ou sobreposição de sons, com apoio do *software* Movavi Video Editor (https://www.movavi.com/pt/videoeditor/). Após a verificação individual dos fragmentos foi realizada a extração dos sons, no formato *Waveform Audio Format* (wav). Os arquivos de áudio resultantes foram submetidos a um filtro digital para eliminar possíveis ruídos, visando assegurar vocalizações isoladas para análises

alinhadas às características acústicas das vocalizações. A aplicação do filtro pode ser feita em tempo de análise/execução, podendo ser usado no processo de coleta relacionado às aplicações reais do sistema.

As aproximadamente 1000 horas de filmagens durante o confinamento geraram 357 vocalizações individuais dos animais. Dado o contexto de controle, esses sons foram categorizados como vocalizações normais. A análise dos registros de vídeo dos manejos foi realizada da mesma forma, gerando 186 vocalizações individuais. Para manter a consistência dos dados, o comprimento de duração dos arquivos de áudio foram padronizados para 3 segundos, o que também foi aplicado às vocalizações capturadas em ambiente de confinamento. Assim como nas vocalizações em confinamento, as vocalizações em manejo também foram submetidas a uma etapa de filtragem de ruídos, visando garantir a qualidade das amostras de áudio. Dada a natureza desse contexto, caracterizado por interações intensas entre animais e humanos, esses sons foram rotulados como vocalizações de estresse.

3.3 Análise acústica

Para compreender e analisar as características das vocalizações coletadas, foram feitas análises nos domínios da amplitude e frequência. Para esse propósito, empregou-se um software em linguagem Python, utilizando as bibliotecas Librosa (https://librosa.org/), Matplotlib (https://matplotlib.org/) e Parselmouth (https://pypi.org/project/praat-parselmouth/) com a finalidade de extrair e visualizar os sinais de áudio nos domínios do tempo e frequência, bem como a sua representação na forma de onda. A Fig. 4 apresenta duas vocalizações durante os períodos de confinamento e manejo nos domínios do tempo e da frequência.

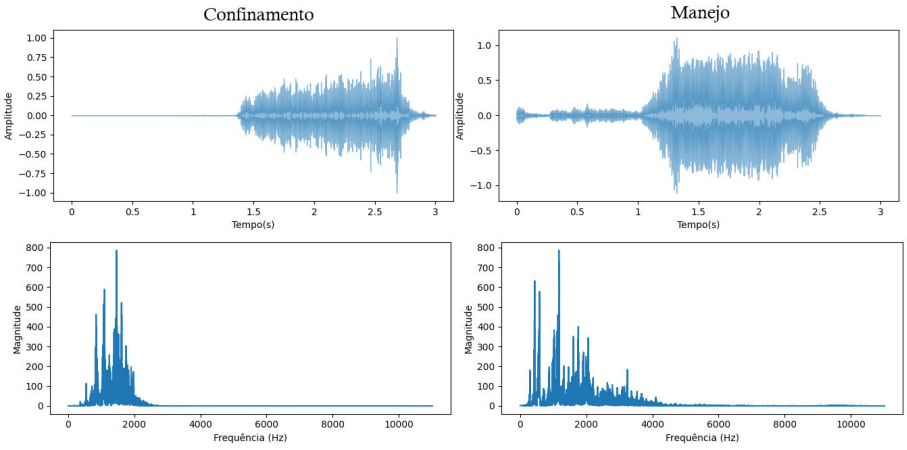


Figura 4. Comparação de vocalizações nos domínios do tempo e frequência

No âmbito da frequência, foram examinadas a densidade espectral de energia (*Power Spectral Density* - PSD) e o espectrograma das vocalizações, buscando compreender a distribuição das principais frequências e identificar eventuais disparidades entre elas. A Fig. 5 exibe gráficos comparativos da análise no domínio da frequência.

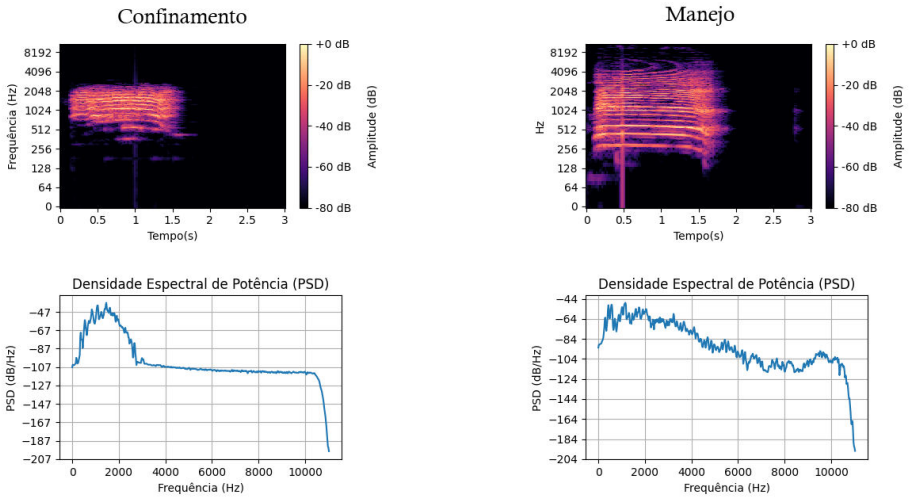


Figura 5. Comparação de vocalizações através do espectrograma e PSD

A análise visual revelou diferenças entre as vocalizações de confinamento e manejo. Em termos de frequência, as representações visuais mostram que as vocalizações durante momentos de estresse apresentavam frequências mais elevadas, evidenciadas claramente nos espectrogramas e nos gráficos de PSD. A análise visual foi importante para destacar indícios sobre mudança de produção vocal em diferentes contextos, incentivando investigações mais aprofundadas sobre as características acústicas das vocalizações. Após a análise visual das vocalizações, investigou-se os principais parâmetros acústicos descritos na literatura sobre as vocalizações de bovinos, incluindo a frequência fundamental (F0), os formantes F1 - F4, *jitter*, *shimmer*, harmonia e intensidade das vocalizações [35,38,25,14,23,18,19].

A frequência fundamental, também chamada de *pitch* ou F0, é a componente de frequência mais baixa de um sinal sonoro, que se relaciona harmonicamente com as outras parciais, o que significa que a frequência da maioria das parciais está relacionada à frequência da parcial mais baixa por uma pequena proporção de números inteiros [9]. Em estudos sobre vocalizações animais, como os de bovinos, o *pitch* pode ser uma métrica relevante para avaliar variações na comunicação vocal, refletindo diferenças na emoção, comportamento ou estado de saúde dos animais [35,11].

Os formantes do espectro podem ser definidos como picos de energia em uma região do espectro sonoro. São caracterizados pela frequência do pico, pelo fator de ressonância e pelo nível de amplitude relativa do som [19]. Um formante é um modo natural de vibração (ressonância) do trato vocal, e como a maioria dos mamíferos não consegue alterar a forma ou as dimensões do seu trato vocal, pois ele é fisicamente restringido por estruturas esqueléticas, o comprimento do trato vocal em bovinos correlaciona-se significativamente com a dispersão de formantes [11].

Em bovinos, a análise dos formantes do espectro pode fornecer dicas e indícios sobre a idade, tamanho e/ou gênero do vocalizador [35]. Mesmo para animais que podem alterar sua voz e formato do trato, o formante mínimo e a dispersão ainda se correlacionam com a idade e o tamanho corporal [11]. Em bezerros de corte, foi demonstrado que as frequências dos formantes diminuíram à medida que os bezerros envelheceram, sendo um resultado direto do crescimento e desenvolvimento dos animais [35].

O *jitter* e o *shimmer* representam variações na frequência fundamental. Enquanto o *jitter* indica a variabilidade ou perturbação na frequência fundamental, o *shimmer* refere-se à mesma perturbação, mas relacionada à amplitude da onda sonora, ou seja, à intensidade da emissão vocal. O *jitter* é influenciado principalmente pela falta de controle de vibração das pregas vocais, enquanto o *shimmer* está relacionado à redução da resistência glótica e lesões de massa nas pregas vocais, frequentemente correlacionado com a presença de ruído. Em estudos com bovinos, valores elevados para *jitter* e *shimmer* foram associados à presença de estresse [37].

A harmonia representa o grau de periodicidade acústica, sendo medida como a razão entre a energia harmônica e a energia não harmônica na vocalização. Valores mais altos refletem vocalizações mais tonais [11]. Em estudos com animais, a harmonia foi empregada na caracterização e identificação de vocalizações [11,20,21].

A intensidade é uma medida da potência do som por unidade de área, descrevendo a quantidade de energia sonora transmitida por uma onda sonora em uma determinada região. Quanto maior a intensidade, mais energia sonora está presente e, portanto, o som é percebido como mais alto. Em estudos com animais, a intensidade tem sido utilizada para analisar vocalizações em condições estressantes e não estressantes [25,20].

A análise acústica indica diferenças significativas entre vocalizações provenientes de situações de confinamento e de manejo, apontando que é possível usar técnicas de aprendizado de máquina para discernir os dois tipos de sons. Essa análise também permite um ponto de comparação entre os resultados deste estudo e aqueles encontrados na literatura. Para cada uma das características examinadas, foram calculadas a média e o desvio padrão para todos os valores associados a cada tipo de vocalização.

3.4 Aprendizado profundo

O módulo de identificação de estresse animal foi implementado na linguagem Python, com o suporte de bibliotecas especializadas como Librosa (https://librosa.org/), Scikit-Learn (https://scikit-learn.org/), TensorFlow (https://www.tensorflow.org/) e Keras (https://keras.io/).

Com base na revisão da literatura, as técnicas selecionadas para análise e classificação de sons foram o MFCC (*Mel-frequency cepstral coefficients*) e redes neurais convolucionais (CNN). A eficácia dessas abordagens na identificação e classificação de sons animais foi amplamente discutida em diversos trabalhos [3,7,15,22,17,27].

O módulo de classificação foi dividido em duas etapas principais:

- **Extração de características acústicas**: os coeficientes MFCC foram extraídos dos arquivos de áudio – técnica amplamente usada na análise de sons para extração de características no domínio da frequência, especialmente em aplicações de reconhecimento de voz [6].
- **Criação dos modelos de classificação**: redes neurais convolucionais foram treinadas com os dados extraídos na etapa anterior para identificar e classificar sons, permitindo o reconhecimento de padrões específicos.

O MFCC é a representação do *cepstrum* real de um sinal janelado em tempo curto derivado da transformada rápida de Fourier (FFT), em escala de frequências não lineares, denominada escala Mel. A utilização da escala Mel visa simular o comportamento do sistema auditivo humano [34]. A extração das *features* de MFCC é realizada a partir dos seguintes passos: *(i) pre emphasis*; *(ii) framing and windowing*; *(iii)* FFT/DFT; *(iv) mel filter bank*; *(v)* IFFT/DCT. As características MFCC são utilizadas em sistemas de reconhecimento de fala, identificação de locutor, processamento de linguagem natural e sistemas de controle de voz devido à sua capacidade de capturar características discriminativas da fala [34], sendo comumente usadas como entrada para modelos de aprendizado de máquina.

Cada vocalização coletada foi processada individualmente pelo algoritmo MFCC para a extração de suas características acústicas. As informações extraídas, juntamente com a classe correspondente de cada som, foram armazenadas em um *array* dinâmico preenchido de forma incremental à medida que as características MFCC eram obtidas a partir dos arquivos de vocalizações.

O extrator de características MFCC foi implementado com o uso da biblioteca Librosa, que possibilitou a extração de informações acústicas a partir de arquivos de áudio, além da definição de parâmetros como frequência de amostragem e número de coeficientes desejados. No software desenvolvido, a frequência de amostragem foi configurada em 44,1 kHz, o número de coeficientes MFCC em 13, o tamanho da janela para a transformada de Fourier em 2048, e o espaçamento entre amostras em 512. A escolha dessas configurações foi baseada nas características dos sinais de áudio processados, nas convenções adotadas para esses parâmetros na biblioteca Librosa e em referências da literatura, garantindo a eficácia e a precisão do extrator de características [31,39,34,17].

O extrator desenvolvido foi aplicado em 200 das 357 vocalizações de confinamento, enquanto todas as 186 vocalizações de manejo foram utilizadas. Essa definição foi feita para manter um equilíbrio entre as classes de estresse e normal, evitando que os classificadores fossem treinados com classes desbalanceadas, o que poderia fazer os modelos favorecerem a classe majoritária no processo de aprendizagem.

As características MFCC extraídas foram normalizadas com a técnica *z-score* para padronizar os valores dos dados. O *z-score* transforma os valores originais de cada característica, subtraindo a média e dividindo pelo desvio padrão. A Eq. (1) apresenta o cálculo do *z-score*, onde x é o valor individual, μ é a média das amostras e σ é o desvio padrão das amostras.

$$z - score = \frac{x - \mu}{\sigma} \tag{1}$$

A normalização dos dados é essencial para evitar que características com escalas distintas influenciem negativamente o treinamento dos modelos, promovendo maior estabilidade e uma convergência mais eficiente. Após a extração das características e o tratamento dos dados, o array resultante, juntamente com as categorias correspondentes de todas as vocalizações, foi armazenado em um DataFrame que foi utilizado como entrada para os modelos de classificação de estresse animal.

O modelo de aprendizado de máquina usado foi baseado em redes neurais convolucionais (CNN), que emergiram do estudo do córtex visual do cérebro e têm sido utilizadas no reconhecimento de imagens desde os anos 1980. As CNN não estão restritas à percepção visual, sendo também bem-sucedidas em outras tarefas, como reconhecimento de voz ou processamento de linguagem natural [10]. Este trabalho fez uso de três arquiteturas distintas de CNN, com níveis diferentes de complexidade: uma estrutura básica, uma intermediária e uma versão mais robusta. Essa abordagem teve como objetivo avaliar o desempenho das redes sob diferentes perspectivas, investigando se variações na complexidade das arquiteturas influenciam significativamente a eficácia na classificação de vocalizações indicativas de estresse em bovinos.

A arquitetura típica de uma CNN consiste em uma camada de entrada, camadas convolucionais, camadas de agrupamento (ou *pooling*) e camadas totalmente conectadas [33]. As camadas convolucionais aprendem as características de baixo nível, as camadas de *pooling* reduzem o tamanho espacial das características convolucionais, diminuindo assim o custo computacional [33], e as camadas totalmente conectadas aprendem a discernir as classes dos dados [10].

As arquiteturas usadas neste trabalho foram as seguintes:

1. **Arquitetura Básica**:
   - Número de camadas: 5 (1 camada convolucional, 1 camada de *max pooling*, 1 camada *flatten*, 1 camada totalmente conectada, 1 camada de saída).
   - Camada convolucional: 64 filtros de tamanho 3 x 3, ativação leaky ReLU.
   - *Pooling*: Max *pooling* de tamanho 2 x 2.

- Camada totalmente conectada: 32 neurônios, ativação leaky ReLU.
- Camada de saída: 1 neurônio, ativação sigmoid.

2. **Arquitetura Intermediária**:
   - Número de camadas: 7 (2 camadas convolucionais, 2 camadas de *max pooling*, 1 camada *flatten*, 1 camada totalmente conectada, 1 camada de saída).
   - Camadas convolucionais: 2 camadas, 64 filtros de tamanho 3 x 3 e 32 filtros de tamanho 3 x 3, ativação leaky ReLU.
   - *Pooling*: Max *pooling* de tamanho 2 x 2 após cada convolução.
   - Camada totalmente conectada: 64 neurônios, ativação leaky ReLU.
   - Camada de saída: 1 neurônio, ativação sigmoid.
3. **Arquitetura Robusta**:
   - Número de camadas: 10 (3 camadas convolucionais, 3 camadas de *max pooling*, 1 camada *flatten*, 2 camada totalmente conectada, 1 camada de saída).
   - Camadas convolucionais: 3 camadas, 256 filtros de tamanho 3 x 3, 128 filtros de tamanho 3 x 3 e 64 filtros de tamanho 3 x3, ativação leaky ReLU.
   - *Pooling*: Max *pooling* de tamanho 2 x 2 após cada convolução.
   - Camadas totalmente conectadas: 2 camadas, 128 neurônios e 64 neurônios, ativação leaky ReLU.
   - Camada de saída: 1 neurônio, ativação sigmoid.

A Arquitetura Básica consiste em uma estrutura simples de rede convolucional, composta por uma única camada convolucional seguida por uma camada de *pooling* e uma camada totalmente conectada. Seu objetivo principal é oferecer uma abordagem direta para classificar as vocalizações de estresse em bovinos, focando em capturar características fundamentais das vocalizações. Suas vantagens incluem simplicidade e eficiência computacional, fácil de entender e rápida de treinar. No entanto, sua capacidade de aprendizado pode ser limitada devido à falta de camadas adicionais para extrair representações mais complexas.

A Arquitetura Intermediária foi projetada para superar as limitações da arquitetura básica, incorporando camadas convolucionais adicionais. Com duas camadas convolucionais e duas camadas de *pooling*. Esta arquitetura visa melhorar a capacidade de representação da rede, permitindo que aprenda características mais abstratas das vocalizações. Suas vantagens incluem uma maior capacidade de aprendizado e generalização devido à inclusão de camadas adicionais.

A Arquitetura Robusta apresenta um maior número de camadas convolucionais e totalmente conectadas. Com três camadas convolucionais, três camadas de pooling e duas camadas totalmente conectadas, esta arquitetura visa capturar representações ainda mais detalhadas e abstratas das vocalizações. Seus pontos fortes incluem uma capacidade de aprendizado superior e uma melhor capacidade de generalização devido à sua profundidade e complexidade. No entanto, sua maior complexidade também pode tornar o treinamento mais demorado e exigir recursos computacionais, além de aumentar o risco de *overfitting*.

Na configuração dos parâmetros para o treinamento de redes neurais, uma série de escolhas precisam ser feitas para otimizar o desempenho do modelo. Esses parâmetros incluem a escolha do otimizador, a taxa de aprendizagem, o método de inicialização de pesos, número de épocas, entre outros. A seleção desses parâmetros influencia significativamente a convergência do modelo, sua capacidade de generalização e a eficácia na resolução do problema em questão. Este processo de ajuste fino dos parâmetros visa encontrar a combinação mais adequada que maximize o desempenho da rede neural para a tarefa específica em análise. Após testes de adequação, s parâmetros de treinamento foram definidos como :

- Otimizador: Adam
- Inicialização de pesos: Glorot
- Taxa de aprendizagem: 0,01
- Épocas: 200
- *Batch size*: 32

Em todas as arquiteturas implementadas foram empregadas técnicas de regularização como *dropout* e *batch normalization*, com o objetivo de mitigar o sobreajuste dos modelos. A taxa de aprendizagem foi fixada em 0,01 e ajustada dinamicamente durante o treinamento, sendo reduzida pela metade (fator = 0,5) caso não houvesse melhorias no treinamento após 10 épocas. O treinamento das redes neurais foi realizado a partir de cinco repetições de validações cruzadas (*cross-validation*) com k = 5.

4 Resultados

4.1 Análise acústica

A análise acústica foi avaliada estatisticamente para verificar se vocalizações de confinamento e manejo tinham diferenças significativas. O teste t de Student foi aplicado sobre as características acústicas extraídas das vocalizações coletadas. Essa análise permite avaliar se existem diferenças significativas entre as médias de grupos independentes. O objetivo principal da análise estatística é determinar se a variabilidade observada nas médias entre os grupos é maior do que a variabilidade esperada devido ao acaso. A análise estatística foi realizada com base nos parâmetros acústicos de frequência fundamental (F0), formantes do espectro F1-F4, *jitter*, *shimmer*, harmonia e intensidade extraídos das vocalizações de dois grupos independentes:

O teste de hipóteses foi realizado para cada parâmetro definido, comparando-se as médias obtidas para cada grupo independente. O teste t de Student foi empregado para avaliar se as médias observadas dos grupos de confinamento e manejo apresentaram diferenças estatisticamente significativas ou são simplesmente variações aleatórias nos dados, com um nível de confiança de 5%. Sendo assim, as hipóteses levantadas são:

- Hipótese Nula (H_0): não há diferença significativa entre as médias dos parâmetros acústicos para vocalizações de confinamento e de manejo. Qualquer variação observada é atribuída ao acaso.
- Hipótese Alternativa (H_A): existe uma diferença significativa entre as médias dos parâmetros acústicos para vocalizações de confinamento e de manejo. A variação observada não é devida ao acaso, indicando uma relação verdadeira entre situações de estresse e mudanças nos parâmetros de vocalizações.

A Tabela 1 apresenta os resultados obtidos para o teste de análise de variância para cada um das características acústicas analisadas, onde é possível observar que apenas para F0 max e para a harmonia não houve diferenças significativas entre as vocalizações de confinamento e manejo.

Tabela 1. Análise de variância dos parâmetros vocais por testes de t de Student

Parâmetros	**Confinamento**	**Manejo**	***P-value***
Média F0 (Hz)	172,87 ± 25,27	276,23 ± 47,38	< 0,001***
Min F0 (Hz)	81,74 ± 14,57	132,54 ± 29,24	< 0,001***
Max F0 (Hz)	436,17 ± 57,25	458,70 ± 51,46	0,1133
Média F1 (Hz)	848,34 ± 51,94	940,47 ± 54,83	< 0.001***
Média F2 (Hz)	1445,04 ± 61,42	1582,79 ± 58,74	< 0.001***
Média F3 (Hz)	2029,96 ± 89,06	2421,29 ± 103,72	< 0.001***
Média F4 (Hz)	3435,90 ± 93,31	3802,71 ± 92,55	0,004**
Jitter (%)	1,61 ± 0,63	3,85 ± 0,81	< 0.001***
Shimmer (%)	14,93 ± 1,7	19,08 ± 1,8	< 0.001***
Harmonia (dB)	7,12 ± 0,17	7,66 ± 0,18	0,198
Intensidade (dB)	48,74 ± 3,35	57,59 ± 4,37	< 0.001***

Os resultados obtidos destacam que, para a maioria dos parâmetros analisados, as vocalizações de animais durante o manejo apresentam médias estatisticamente distintas em comparação com as vocalizações durante o confinamento. Essa disparidade sugere que o contexto do manejo pode influenciar significativamente as características acústicas das vocalizações dos animais, indicando a existência de padrões distintos entre os dois cenários.

Para a frequência fundamental (F0), tanto a média quanto o mínimo mostram diferenças significativas entre os grupos de confinamento e manejo ($p < 0{,}001$), sendo a média maior no grupo de manejo. No entanto, a frequência fundamental máxima (Max F0) não apresentou diferença significativa entre os grupos ($p = 0{,}1133$). Os resultados obtidos indicam uma variação considerável na modulação vocal entre os contextos de confinamento e manejo.

As médias dos primeiros quatro formantes (F1, F2, F3, F4) exibiram diferenças significativas entre os grupos ($p < 0{,}001$), sendo as médias dos formantes no grupo de manejo consistentemente superiores em comparação com o grupo de confinamento. Esses resultados sugerem que as características espectrais da vocalização bovina são sensíveis às condições de manejo, possivelmente refletindo a influência do ambiente na produção vocal dos animais.

Para os parâmetros de variabilidade vocal, *jitter* e *shimmer*, há diferenças significativas entre os grupos ($p < 0{,}001$), ambos sendo maiores no grupo de manejo. Em relação ao parâmetro de *jitter*, constatou-se que os animais em manejo apresentaram vocalizações com índices mais elevados ($3{,}85 \pm 0{,}81\%$) em comparação com o grupo de confinamento ($1{,}61 \pm 0{,}63\%$). Essa diferença na variabilidade temporal das vocalizações pode indicar uma resposta vocal mais instável ou agitada nos bovinos submetidos ao manejo. Analisando o parâmetro *shimmer*, observou-se que o grupo de manejo apresentou um índice de variação maior ($19{,}08 \pm 1{,}8\%$) em comparação com o grupo de confinamento ($14{,}93 \pm 1{,}7\%$). Isso sugere uma maior variação na amplitude das vocalizações nos bovinos sob condições de manejo, indicando possíveis alterações na regularidade e estabilidade vocal.

A medida de harmonia em decibéis não mostrou diferença significativa entre os grupos ($p = 0{,}198$), com médias de $7{,}12 \pm 0{,}17$ (confinamento) e $7{,}66 \pm 0{,}18$ (manejo). Isso sugere que, ao contrário de outros parâmetros, a harmonia vocal não foi afetada de maneira significativa pelas condições de manejo.

No que se refere à intensidade vocal, constatou-se que o grupo de manejo exibiu uma intensidade vocal mais elevada ($57{,}59 \pm 4{,}37$ dB) em comparação com o grupo de confinamento ($48{,}74 \pm 3{,}35$ dB), apresentando diferença significativa entre os grupos ($p < 0{,}001$). Essa diferença sugere uma vocalização mais intensa em situações de manejo, indicando uma resposta vocal mais vigorosa e enérgica nessas condições.

Esses resultados indicam que as condições de manejo têm um impacto significativo nas características vocais, com diferenças significativas em parâmetros como frequência fundamental, formantes, variabilidade vocal e intensidade entre os grupos de confinamento e manejo. Em comparação aos trabalhos encontrados na literatura, é possível observar semelhanças nos resultados. A Tabela 2 sintetiza os principais resultados obtidos na literatura sobre as diferenças nos parâmetros acústicos das vocalizações de bovinos em condições diversas. Ainda que os resultados não possem ser diretamente comparáveis, por diferenças em objetivos e técnicas, indicam similaridades e diferenças entre os resultados encontrados neste trabalho.

Em conformidade com as descobertas da literatura, que destacam diferenças significativas entre animais em condições distintas, a análise acústica revelou que as vocalizações de bovinos sob estresse apresentam parâmetros acústicos diferentes daqueles produzidos por bovinos em situações normais, especialmente em relação às propriedades de frequência. Ao comparar os resultados médios obtidos com os de estudos similares, observa-se que os valores estão em faixas próximas, o que sugere que os dados são confiáveis e estão alinhados com o esperado para vocalizações de gado bovino. Essa consistência fortalece a validade dos resultados e a sua contribuição para uma melhor compreensão dos padrões de vocalizações associados ao estresse animal.

A análise acústica permitiu identificar mudanças significativas na produção vocal de bovinos em diferentes contextos. Esses resultados incentivaram investigações mais aprofundadas sobre as características acústicas das vocalizações,

Tabela 2. Comparação entre os resultados obtidos e a literatura

	Presente trabalho	[37]	[35]	[38]	[11]
Média F0 (Hz)	172 - 276	191	81 - 152	214 - 221	183 - 286
Min F0 (Hz)	81 - 132	78	74 - 121	NA	66 - 76
Max F0 (Hz)	436 - 458	352	84 - 198	NA	473 - 559
Média F1 (Hz)	848 - 940	689 - 1015	228 - 391	784 - 832	NA
Média F2 (Hz)	1445 - 1582	1675 - 1942	634 - 1162	1710 - 1570	NA
Média F3 (Hz)	2029 - 2421	3079 - 3412	1064 - 1939	2418 - 2675	NA
Média F4 (Hz)	3435 - 3802	4939 - 5296	1513 - 2722	3818 - 3559	NA
Jitter (%)	1,61 - 3,85	1,09 - 2,77	2 - 4	NA	1 - 5
Shimmer (%)	14 - 19	4,85 - 8,97	15 - 17	NA	9 - 18
Harmonia (dB)	7,12 - 7,66	5,86 - 12,71	NA	NA	7,72 - 11,97
Intensidade (dB)	48 - 57	60 - 87	NA	68 - 71	NA

bem como o uso de redes neurais artificiais para automatizar a identificação de estados de estresse em animais.

4.2 Aprendizado profundo

A análise dos resultados obtidos no treinamento das diferentes arquiteturas de redes CNN foi feita a partir das métricas de acurácia, precisão, revocação e *F1-score*. A Tabela 3 apresenta os valores médios obtidos para cada uma das métricas das três arquiteturas implementadas após a execução do treinamento por repetições de validação cruzada, onde pode-se observar que todas as arquiteturas alcançaram um bom desempenho na classificação das vocalizações. A arquitetura mais robusta apresentou desempenho superior em comparação com as de menor complexidade. No entanto, os ganhos em acurácia, precisão, revocação e *F1-score* foram modestos. Isso sugere que os recursos adicionais de computação investidos não se traduziram em melhorias proporcionais na capacidade de classificação de vocalizações.

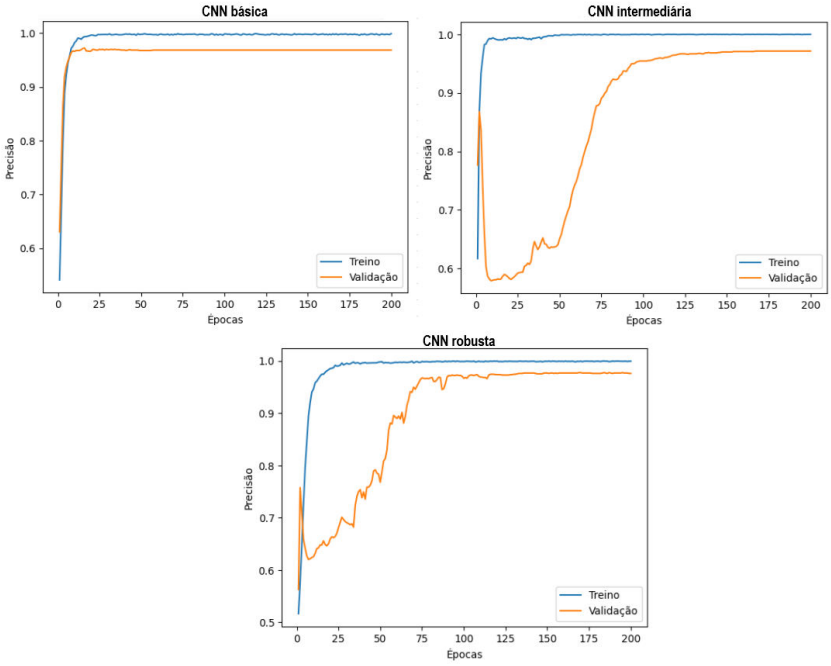
Tabela 3. Médias para as métricas de acurácia, precisão, revocação e *F1-score*

Arquitetura	**Acurácia**	**Precisão**	**Revocação**	***F1-score***
Básica	96,92%	95,37%	98,63%	96,97%
Intermediária	97,88%	96,87%	98,95%	97,90%
Robusta	98,74%	98,37%	99,11%	98,74%

Também observa-se que as arquiteturas CNN apresentaram melhores desempenhos na classificação das vocalizações de estresse em comparação com as vocalizações normais, evidenciado pelos maiores valores de revocação. Esses resultados ressaltam a eficácia das arquiteturas CNN na tarefa de classificação de vocalizações animais, comprovando seu potencial para aplicações práticas em estudos comportamentais e de bem-estar animal.

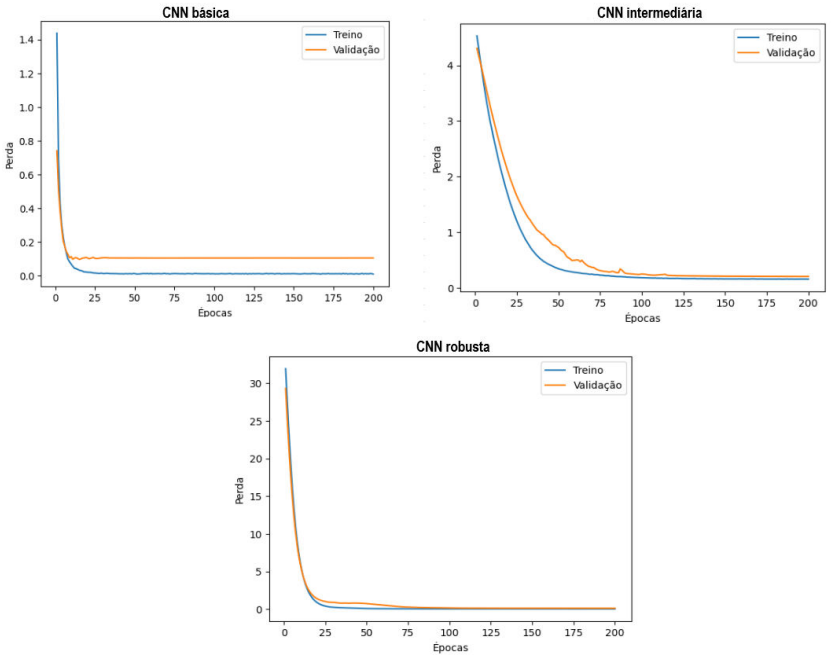
A Fig. 6 apresenta os gráficos de precisão durante o treinamento das três arquiteturas CNN, onde observa-se comportamentos distintos entre as arquiteturas. A arquitetura básica alcançou estabilidade rapidamente, por volta da época 20, mantendo-se constante até o término do treinamento. Já a arquitetura intermediária teve um início de treinamento com perda significativa de precisão, que gradualmente se recuperou até atingir um platô por volta da época 100, mantendo-se estável até o final. Por sua vez, a arquitetura robusta também apresentou uma queda inicial na precisão, embora menos acentuada que a intermediária. Posteriormente, observou-se alguma variação, mas ela alcançou um platô por volta da época 100, mantendo-se praticamente constante até o término do treinamento.

Figura 6. Precisões nos treinamentos das arquiteturas CNN



A Fig. 7 apresenta os gráficos de perda durante os treinamentos das três arquiteturas CNN. Na arquitetura básica, observa-se que a perda se estabiliza rapidamente no início do treinamento, mantendo-se constante até o final. Já na arquitetura intermediária, a curva de perda é mais suave, alcançando o platô por volta da época 100. Por sua vez, a arquitetura robusta também atinge a estabilidade logo no início do treinamento, no entanto, diferencia-se da básica pela proximidade entre as perdas de treinamento e validação, enquanto na básica essa diferença é mais significativa.

Figura 7. Perdas nos treinamentos das arquiteturas CNN



Para avaliar o desempenho das arquiteturas implementadas e determinar se as diferenças observadas nos resultados entre as arquiteturas foram devidas às variações nas complexidades das arquiteturas ou apenas ao acaso, foi realizada uma análise de variância (ANOVA).

O teste de hipóteses foi conduzido utilizando o *F1-score* como a métrica de análise, devido à sua robustez em comparação com a acurácia, precisão e revocação. Foram comparadas as médias do *F1-score* entre as arquiteturas CNN. As hipóteses formuladas foram as seguintes:

- Hipótese nula (H_0): Não há diferença significativa entre as médias do *F1-score* entre as arquiteturas. Qualquer variação observada é atribuída ao acaso.
- Hipótese alternativa (H_A): Existe uma diferença estatisticamente significativa entre as médias do *F1-score* entre as arquiteturas. A variação observada não é devida ao acaso, indicando uma relação genuína entre a complexidade das redes e o desempenho na classificação de vocalizações normais e de estresse.

A Tabela 4 apresenta os resultados da ANOVA, para um nível de confiança de 5%, para os três modelos de arquiteturas CNN implementadas.

A análise comparativa entre os modelos de redes neurais revelou diferenças significativas em relação ao desempenho das arquiteturas, conforme evidenciado

Tabela 4. Análise de variância por testes de Tukey entre os modelos de CNN

Arquitetura vs Arquitetura	**F1-score**		**Diferença**	***P-value***
CNN básica vs CNN intermediária	96,97	97,90	0,93	0,158
CNN básica vs CNN robusta	96,97	98,74	1,77	0,002**
CNN intermediária vs CNN robusta	97,90	98,74	0,84	0,22

pelos testes de Tukey, apontando uma melhora estatisticamente significativa entre as arquiteturas básica e robusta (*P-value* = 0,002). Contudo, entre as arquiteturas básica e intermediária, e intermediária e robusta não há evidências estatísticas de que o aumento de complexidade refletiu em melhores resultados. Esses resultados destacam a influência da arquitetura e da complexidade do modelo no desempenho das redes neurais, sugerindo que, em determinados contextos, aumentar a complexidade do modelo pode resultar em melhorias significativas na capacidade de generalização e aprendizado. No entanto, para outras arquiteturas, o aumento de complexidade pode não resultar em maior poder de classificação [26].

Realizando um comparativo com os resultados encontrados na literatura, é possível destacar a utilização de redes neurais do tipo CNN para classificação de vocalizações animais em diferentes trabalhos [27,31,30,36,17]. A Tabela 5 apresenta um comparativo entre os resultados obtidos por este trabalho e a literatura.

Tabela 5. Comparação entre os resultados obtidos e a literatura

Trabalho	**Característica acústica**	**Resultados**
Presente trabalho	MFCC	Acurácia = 98,74% Precisão = 98,37% Revocação = 99,11% F1-score = 98,74%
[31]	MFCC	Acurácia = 84%
[17]	MFCC	Acurácia = 94,18%
[30]	MFCC	Acurácia = 75%
[27]	Espectrograma	F1-score = 70,90%
[32]	Espectrograma	Acurácia = 96,2%
[36]	Espectrograma	F1-score = 61,7%
[8]	23 parâmetros vocais	F1-score = 89,4%

Ao analisar a Tabela 5, destaca-se que a maioria dos estudos encontrados na literatura empregou os coeficientes MFCC como características de análise acústica e também observa-se o uso do espectrograma como uma característica de interesse para o modelo de classificação, como também o uso de parâmetros vocais nos domínios da frequência e amplitude. Ainda que os resultados numéricos não possam ser comparados, por diferenças nos dados e métodos usados, além de diferenças entre as raças de bovinos analisados, os resultados obtidos

neste trabalho são consistentes com estudos anteriores que adotaram abordagens semelhantes. Os resultados alcançados neste trabalho não foram testados com bases independentes porque não conseguimos encontrar dados abertos rotulados na forma deste trabalho. Ainda assim, os resultados obtidos são promissores.

Os resultados deste trabalho revelaram a viabilidade em empregar as características extraídas do MFCC como base para o treinamento de redes neurais na identificação de estresse em bovinos. No estudo de arquiteturas CNN para a classificação de vocalizações, todas alcançaram bons resultados. Esse fato pode ser um indício para a preferência predominante na literatura pelo uso de redes CNN na análise acústica de vocalizações animais.

Entre as diferentes complexidades de arquiteturas, a variante robusta apresentou resultado estatisticamente superior à variante básica. Contudo, é importante ressaltar que o aumento na complexidade vem acompanhado de maiores exigências computacionais, como requisitos de memória e poder de processamento. As redes robustas demandaram consideravelmente mais tempo de treinamento, além de exigir maior poder de processamento em comparação com as redes menos complexas. Portanto, ao escolher uma arquitetura de rede neural é importante avaliar os aspectos computacionais, principalmente em sistemas embarcados e de tempo real, onde as limitações de *hardware* podem ser restritivas para o uso eficaz de redes neurais mais robustas.

5 Conclusão

A pecuária bovina de corte é uma das principais fontes de renda no Brasil. No entanto, enfrenta desafios para aprimorar sua produtividade, sendo um ponto crucial a crescente exigência dos consumidores e dos países exportadores quanto à qualidade da carne bovina e à garantia de qualidade de vida dos animais, demandando produtos de maior qualidade.

Nesse sentido, há um foco de estudos que objetivam analisar e identificar o estresse de animais. A análise acústica surge como uma das principais formas de estudar o bem-estar animal, uma vez em que precários níveis de bem-estar estão fortemente relacionados com a perda da qualidade do produto final. Assim, esse trabalho teve como objetivo a análise de vocalizações em momentos psicologicamente distintos para os animais, com a finalidade de identificar mudanças nos parâmetros vocais dos animais quando submetidos a condições estressantes.

Os resultados encontrados pela análise acústica, em concordância com estudos na literatura, revelaram que bovinos submetidos a condições psicologicamente estressantes tendem a produzir vocalizações com parâmetros significativamente diferentes daqueles em condições não estressantes.

Para o treinamento de redes neurais, os resultados obtidos alcançaram acurácias variando de 96,92% a 98,74%, atestou-se a eficácia do MFCC em capturar as características essenciais das vocalizações, possibilitando a distinção entre momentos de estresse e não estresse. Os resultados da pesquisa também confirmaram as hipóteses levantadas de que os sons emitidos por animais estressados

são distintos dos emitidos por animais tranquilos, e de que redes neurais artificiais são capazes de discernir situações de estresse e não estresse.

Como trabalhos futuros, busca-se avançar na pesquisa com a expansão da base de dados com a inclusão de vocalizações em condições adicionais além do estudo de uma maior variedade de parâmetros acústicos. Nossas bases de dados também serão tornadas públicas, dentro dos princípios FAIR de ciência aberta. Essas iniciativas podem aprimorar ainda mais a caracterização de vocalizações de estresse, consolidando assim o avanço no desenvolvimento de métodos não invasivos para monitoramento do bem-estar animal na indústria pecuária.

A pesquisa possui potencial para contribuir com o avanço do conhecimento sobre as características de vocalizações animais em condições normais e de estresse, enriquecendo o estado atual do conhecimento nesse campo de estudo. Além disso, pode servir de estímulo para que novos trabalhos investiguem o estresse animal por meio das vocalizações emitidas, agregando ao estado da arte nesse campo de estudo, além de consolidar o avanço no desenvolvimento de métodos não invasivos para monitoramento do bem-estar animal na indústria pecuária bovina.

Referências

1. Arksey, H., O'Malley, L.: Scoping studies: Towards a methodological framework. Int. J. Social Research Methodology **8**(1), 19–32 (2005)
2. Chelotti, J.O., Vanrell, S.R., Milone, D.H., Utsumi, S.A., Galli, J.R., Rufiner, H.L., Giovanini, L.L.: A real-time algorithm for acoustic monitoring of ingestive behavior of grazing cattle. Computers and Electronics in Agriculture **127**, 64–75 (2016)
3. Chung, Y., Lee, J., Oh, S., Park, D., Chang, H., Kim, S.: Automatic detection of cow's oestrus in audio surveillance system. Asian-Australasian journal of animal sciences **26**(7), 1030 (2013)
4. Clapham, W.M., Fedders, J.M., Beeman, K., Neel, J.P.: Acoustic monitoring system to quantify ingestive behavior of free-grazing cattle. Computers and Electronics in Agriculture **76**(1), 96–104 (2011)
5. da Costa, P., JR, M.: Ambiência na produção de bovinos de corte a pasto. Anais de Etologia **18**, 26–42 (2000)
6. Dave, N.: Feature extraction methods LPC, PLP and MFCC in speech recognition. International Journal for Advance Research in Engineering and Technology **1**(6), 1–4 (2013)
7. Deshmukh, O., Rajput, N., Singh, Y., Lathwal, S.: Vocalization patterns of dairy animals to detect animal state. In: Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). pp. 254–257. IEEE (2012)
8. Gavojdian, D., Lazebnik, T., Mincu, M., Oren, A., Nicolae, I., Zamansky, A.: Bovinetalk: Machine learning for vocalization analysis of dairy cattle under negative affective states. arXiv preprint arXiv:2307.13994 (2023)
9. Gerhard, D., et al.: Pitch extraction and fundamental frequency: History and current techniques. Department of Computer Science, University of Regina Regina, SK, Canada (2003)
10. Géron, A.: Hands-on machine learning with scikit-learn. Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems **1** (2019)

11. Green, A.C.: Decoding Holstein-Friesian dairy cattle vocalisations: Applications for welfare assessment. Ph.D. thesis, University of Sydney (2020)
12. Hopkins, D.L., Bruce, H., Li, D.: Final report–causes and contributing factors to dark cutting: Current trends and future directions (2016)
13. IBGE: Indicadores da produção pecuária (2018)
14. Ikeda, Y., Ishii, Y.: Recognition of two psychological conditions of a single cow by her voice. Computers and Electronics in Agriculture **62**(1), 67–72 (2008)
15. Jahns, G.: Call recognition to identify cow conditions—a call-recogniser translating calls to text. Computers and Electronics in Agriculture **62**(1), 54–58 (2008)
16. Jorquera-Chavez, M., Fuentes, S., Dunshea, F.R., Jongman, E.C., Warner, R.D.: Computer vision and remote sensing to assess physiological responses of cattle to pre-slaughter stress, and its impact on beef quality: A review. Meat Science **156**, 11–22 (2019)
17. Jung, D.H., Kim, N.Y., Moon, S.H., Jhin, C., Kim, H.J., Yang, J.S., Kim, H.S., Lee, T.S., Lee, J.Y., Park, S.H.: Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering. Animals **11**(2), 357 (2021)
18. Lee, J., Noh, B., Jang, S., Park, D., Chung, Y., Chang, H.H.: Stress detection and classification of laying hens by sound analysis. Asian-Australasian Journal of Animal Sciences **28**(4), 592 (2015)
19. Lee, J., Zuo, S., Chung, Y., Park, D., Chang, H.H., Kim, S.: Formant-based acoustic features for cow's estrus detection in audio surveillance system. In: 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 236–240. IEEE (2014)
20. Linhart, P., Ratcliffe, V.F., Reby, D., Špinka, M.: Expression of emotional arousal in two different piglet call types. PloS one **10**(8), e0135414 (2015)
21. Maigrot, A.L., Hillmann, E., Briefer, E.F.: Encoding of emotional valence in wild boar (sus scrofa) calls. Animals **8**(6), 85 (2018)
22. Manteuffel, G., Schön, P.C.: Measuring pig welfare by automatic monitoring of stress calls. Agrartechische Berichte **29**(1) (2002)
23. Meen, G., Schellekens, M., Slegers, M., Leenders, N., Van Erp-van der Kooij, E., Noldus, L.P.: Sound analysis in dairy cattle vocalisation as a potential welfare monitor. Computers and Electronics in Agriculture **118**, 111–115 (2015)
24. Molento, C.F.M.: Bem-estar e produção animal: aspectos econômicos-revisão. Archives of Veterinary Science **10**(1) (2005)
25. Moura, D., Silva, W., Naas, I., Tolón, Y., Lima, K., Vale, M.: Real time computer stress monitoring of piglets using vocalization analysis. Computers and Electronics in Agriculture **64**(1), 11–18 (2008)
26. Paiva, B.T.: Um estudo sobre modelos de redes neurais para identificação de estresse em vocalizações de gado bovino. Master's thesis, Programa de Pós-graduação em Computação Aplicada, Universidade Federal do Pampa (2024)
27. Pandeya, Y.R., Bhattarai, B., Afzaal, U., Kim, J.B., Lee, J.: A monophonic cow sound annotation tool using a semi-automatic method on audio/video data. Livestock Science **256**, 104811 (2022)
28. Polizel Neto, A., Zanco, N., Lolatto, D.C., Moreira, P.S., Dromboski, T.: Perdas econômicas ocasionadas por lesões em carcaças de bovinos abatidos em matadouro-frigorífico do norte de Mato Grosso. Pesquisa Veterinária Brasileira **35**(4), 324–328 (2015)
29. Röttgen, V., Schön, P., Becker, F., Tuchscherer, A., Wrenzycki, C., Düpjan, S., Puppe, B.: Automatic recording of individual oestrus vocalisation in group-housed dairy cattle: development of a cattle call monitor. Animal **14**(1), 198–205 (2020)

30. Şaşmaz, E., Tek, F.B.: Animal sound classification using a convolutional neural network. In: 2018 3rd International Conference on Computer Science and Engineering (UBMK). pp. 625–629. IEEE (2018)
31. Sattar, F.: A context-aware method-based cattle vocal classification for livestock monitoring in smart farm. Chemistry Proceedings **10**(1) (2022). https://doi.org/10.3390/IOCAG2022-12233
32. Shorten, P.: Acoustic sensors for detecting cow behaviour. Smart Agricultural Technology **3**, 100071 (2023). https://doi.org/https://doi.org/10.1016/j.atech.2022.100071
33. Silaparasetty, V.: Deep Learning Projects Using TensorFlow 2. Springer (2020)
34. Tiwari, V.: MFCC and its applications in speaker recognition. International Journal on Emerging Technologies **1**(1), 19–22 (2010)
35. de la Torre, M.P., Briefer, E.F., Reader, T., McElligott, A.G.: Acoustic analysis of cattle (bos taurus) mother–offspring contact calls from a source–filter theory perspective. Applied Animal Behaviour Science **163**, 58–68 (2015). https://doi.org/https://doi.org/10.1016/j.applanim.2014.11.017
36. Vidana-Vila, E., Malé, J., Freixes, M., Solıs-Cifré, M., Jiménez, M., Larrondo, C., Guevara, R., Miranda, J., Duboc, L., Mainau, E., et al.: Automatic detection of cow vocalizations using convolutional neural networks. Tampere, Finland (2023)
37. Yajuvendra, S., Lathwal, S.S., Rajput, N., Raja, T.V., Gupta, A.K., Mohanty, T.K., Ruhil, A.P., Chakravarty, A.K., Sharma, P.C., Sharma, V., et al.: Effective and accurate discrimination of individual dairy cattle through acoustic sensing. Applied Animal Behaviour Science **146**(1-4), 11–18 (2013)
38. Yeon, S.C., Jeon, J.H., Houpt, K.A., Chang, H.H., Lee, H.C., Lee, H.J.: Acoustic features of vocalizations of korean native cows (bos taurus coreanea) in two different conditions. Applied animal behaviour science **101**(1-2), 1–9 (2006)
39. Zheng, F., Zhang, G., Song, Z.: Comparison of different implementations of MFCC. Journal of Computer Science and Technology **16**, 582–589 (2001)