



UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
DEPARTAMENTO DE ESTATÍSTICA E INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA APLICADA

ANDERSON DA SILVA SANTOS

CLASSIFICAÇÃO DO COMPORTAMENTO DE
BOVINOS POR MEIO DE SENSORES INERCIAIS:
UMA ABORDAGEM COM APRENDIZAGEM
PROFUNDA

RECIFE – PE

2022

ANDERSON DA SILVA SANTOS

**CLASSIFICAÇÃO DO COMPORTAMENTO DE
BOVINOS POR MEIO DE SENSORES INERCIAIS:
UMA ABORDAGEM COM APRENDIZAGEM
PROFUNDA**

Dissertação submetida à Coordenação do Programa de Pós-Graduação em Informática Aplicada do Departamento de Estatística e Informática - DEINFO - Universidade Federal Rural de Pernambuco, como parte dos requisitos necessários para obtenção do grau de Mestre.

ORIENTADOR: Prof. Dr. Glauco Estácio Gonçalves

CO-ORIENTADOR: Prof. Dr. Victor Wanderley Costa de Medeiros

RECIFE – PE

2022

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

S237c

Santos, Anderson da Silva

CLASSIFICAÇÃO DO COMPORTAMENTO DE BOVINOS POR MEIO DE SENSORES INERCIAIS:
UMA ABORDAGEM COM APRENDIZAGEM PROFUNDA / Anderson da Silva Santos. - 2022.
87 f. : il.

Orientador: Glauco Estacio Goncalves.

Coorientador: Victor Wanderley Costa de Medeiros.

Inclui referências e apêndice(s).

Dissertação (Mestrado) - Universidade Federal Rural de Pernambuco, Programa de Pós-Graduação em
Informática Aplicada, Recife, 2022.

1. Aprendizagem de máquina. 2. Monitoramento de bovinos. 3. Pecuária de precisão. I. Goncalves,
Glauco Estacio, orient. II. Medeiros, Victor Wanderley Costa de, coorient. III. Título

CDD 004

ANDERSON DA SILVA SANTOS

CLASSIFICAÇÃO DO COMPORTAMENTO DE
BOVINOS POR MEIO DE SENSORES INERCIAIS:
UMA ABORDAGEM COM APRENDIZAGEM
PROFUNDA

Aprovada em: 24 de 02 de 2022.

BANCA EXAMINADORA

Prof. Dr. Glauco Estácio Gonçalves (Orientador)
Universidade Federal do Pará
Instituto de Tecnologia

Prof. Dr. Valmir Macário Filho
Universidade Federal Rural de Pernambuco
Departamento de Estatística e Informática

Prof. Dr. Gledson Luiz Pontes de Almeida
Universidade Federal Rural de Pernambuco
Departamento de Engenharia Agrícola

Eu dedico este trabalho à minha família, em especial, à minha irmã Andrielly e ao meus pais Maria e José.

Agradecimentos

Primeiramente, a Deus por toda a sua misericórdia sobre a minha vida, por ter me ajudado em cada etapa durante essa jornada e por ter me concedido saúde, principalmente, nesses tempos difíceis em que vivemos.

A minha família que sempre me apoiou nos estudos, mesmo nos momentos mais difíceis que passamos, por todo o carinho, compreensão e incentivo. Em especial, a minha mãe Maria Aparecida por ser um exemplo de vida para mim tanto na perspectiva pessoal quanto na profissional.

A todos os meus amigos pelo entendimento de algumas ocasiões que não pude estar presente.

Aos colegas de trabalho Luiz Sérgio e Benilton por todo o suporte e amizade. Muito obrigado.

Aos professores Dr. Glauco Gonçalves e Dr. Victor Medeiros por toda paciência, humildade, empenho e bom humor nas orientações e no compartilhamento dos conhecimentos.

Aos meus colegas do Juá Labs pela parceria, ajuda e boas conversas durante os nossos encontros nas quintas-feiras e em outros momentos. Em especial, ao meu amigo de longa data Me. Diego Bezerra pelos conselhos e tantos outros auxílios nas dúvidas até mesmo antes de entrar no mestrado. Foi um prazer revê-lo.

Suas crenças se tornam seus pensamentos.

Seus pensamentos se tornam suas palavras.

Suas palavras se tornam suas ações.

Suas ações se tornam seus hábitos.

Seus hábitos se tornam seus valores.

Seus valores se tornam seu destino.

(Mahatma Gandhi)

Resumo

A pecuária de precisão propõe transformações nos processos de produção por meio de informações precisas, individuais e, em certos casos, em tempo real. Os sistemas de monitoramento do comportamento de bovinos, em especial, são ferramentas que fornecem essas informações, as quais contribuem para o controle de alguns aspectos, tais como o bem-estar e a saúde animal. Em geral, esses sistemas são baseados em dados gerados a partir de sensores embarcados em um acessório preso ao próprio corpo do animal. Diante disso, uma modelagem matemática para a classificação do comportamento de bovinos a partir desses dados é necessária. Na literatura, o uso de algoritmos de aprendizagem de máquina é recorrente e há pouco uso de técnicas de aprendizagem profunda. Nesta dissertação, como representantes da aprendizagem profunda, são exploradas arquiteturas da *Convolutional Neural Network* (CNN) e realizada uma comparação com os cinco classificadores de aprendizagem de máquina mais utilizados para a classificação do comportamento de bovinos. Devido ao forte desbalanceamento na base de dados, foi necessário a aplicação de técnicas de reamostragem. Os resultados mostraram o potencial das CNNs para a classificação do comportamento de bovinos, além do mais, as técnicas de reamostragem tiveram um papel significativo para a melhoria da performance dos modelos.

Palavras-chave: aprendizagem de máquina, monitoramento de bovinos, pecuária de precisão

Abstract

Precision livestock proposes transformations in production processes through precise, individual information and, in certain cases, in real-time. Cattle behavior monitoring systems, in particular, are tools that provide this information, which contribute to the control of some aspects, such as animal welfare and health. In general, these systems are based on data generated from sensors embedded in an accessory attached to the animal's own body. Given this, mathematical modeling for the classification of cattle behavior based on these data is necessary. In the literature, the use of machine learning algorithms is recurrent and there is little use of deep learning techniques. In this dissertation, as representatives of deep learning, architectures of the CNN are explored and a comparison is made with the five most used machine learning classifiers for classifying the behavior of cattle. Due to the strong imbalance in the database, it was necessary to apply resampling techniques. The results showed the potential of CNNs for the classification of cattle behavior, in addition, the resampling techniques played a significant role in improving the performance of the models.

Keywords: livestock monitoring, machine learning, precision livestock

Lista de Figuras

Figura 1 – Exportações Brasileiras do setor de carnes em 2021	17
Figura 2 – Fluxograma de um sistema de classificação tradicional de aprendizagem de máquina	24
Figura 3 – Estrutura da matriz de confusão	30
Figura 4 – Coleira equipada com a placa junto com os sensores	48
Figura 5 – Animal vestido com a coleira de acordo com o padrão	49
Figura 6 – Componentes utilizados na solução de monitoramento (a) Arduino Fio (b) OpenLog (c) Vênus GPS (d) MinIMU-9 v3	49
Figura 7 – Classes dos comportamentos de bovinos levadas em consideração na rotulação: (A) Pastando, (B) Andando, (C) Em pé e (D) Deitado	51
Figura 8 – Fluxograma do método	53
Figura 9 – Distribuição dos dados em relação às classes de comportamentos (a) antes da reamostragem e (b) após a reamostragem	54
Figura 10 – Vetor de características originário	55
Figura 11 – Filtro de convolução com dimensão 3x3 sobre o vetor de características original	56
Figura 12 – Formação do vetor de características por combinação	57
Figura 13 – Arquitetura CNN por combinação	58
Figura 14 – Arquitetura CNN por matriz de ativação	59
Figura 15 – Ativações na matriz de neurônios para padrões de classes diferentes	60
Figura 16 – Fluxograma para a escolha dos hiperparâmetros da CNN	61
Figura 17 – Arquiteturas iniciais das CNNs (a) combinação e (b) matriz de ativação	61
Figura 18 – Performances das CNNs em relação à base de validação no cenário com a base original	63
Figura 19 – Performances das CNNs otimizadas em relação à base de validação no cenário com a base original	64

Lista de Tabelas

Tabela 1 – Informações dos animais utilizados nos experimentos	50
Tabela 2 – Estrutura da base de dados	52
Tabela 3 – Medidas de avaliação por classe de comportamento das CNNs em relação à base de teste - cenário com base de treino original	66
Tabela 4 – Medidas de avaliação por classe de comportamento das CNNs em relação à base de teste - cenário com base de treino <i>bootstrap</i>	66
Tabela 5 – Medidas de avaliação por classe de comportamento das CNNs em relação à base de teste - cenário com base de treino SMOTE	67
Tabela 6 – Lista dos principais hiperparâmetros dos classificadores	69
Tabela 7 – Medidas de avaliação dos modelos - base original	70
Tabela 8 – Medidas de avaliação dos modelos - base <i>bootstrap</i>	70
Tabela 9 – Medidas de avaliação dos modelos - base SMOTE	71
Tabela 10 – Matriz confusão da CNN por combinação em relação à base de teste - cenário com a base de treino original	72
Tabela A.1 – Arquitetura CNN por combinação	85
Tabela A.2 – Arquitetura CNN por matriz de ativações	85
Tabela B.1 – Matriz confusão CNN (combinação) em relação à base de teste - cenário com base de treino original	86
Tabela B.2 – Matriz confusão CNN (combinação) em relação à base de teste - cenário com base de treino <i>bootstrap</i>	86
Tabela B.3 – Matriz confusão CNN (combinação) em relação à base de teste - cenário com base de treino <i>smote</i>	86
Tabela C.1 – Matriz confusão CNN (matriz de ativações) em relação à base de teste - cenário com base de treino original	87
Tabela C.2 – Matriz confusão CNN (matriz de ativações) em relação à base de teste - cenário com base de treino <i>bootstrap</i>	87
Tabela C.3 – Matriz confusão CNN (matriz de ativações) em relação à base de teste - cenário com base de treino <i>smote</i>	87

Lista de Algoritmos

1	SMOTE(T, N, k)	28
---	--------------------------	----

Lista de Siglas

ADASYN	<i>Adaptive Synthetic Sampling</i>
BOCP	<i>Bag of Class Posteriors</i>
BOCPO	<i>Bag of Class Posteriors with Ordering</i>
BoF	<i>Bag-of-Words</i>
BoW	<i>Bag-of-Features</i>
BT	<i>Binary Tree</i>
CE	<i>Cross Entropy</i>
CNN	<i>Convolutional Neural Network</i>
DT	<i>Decision Tree</i>
Embrapa	Empresa Brasileira de Pesquisa Agropecuária
FACOM	Faculdade de Computação
FC	<i>Fully Connected</i>
FCM	<i>Fuzzy C Means</i>
FN	Falso Negativo
FP	Falso Positivo
GPS	<i>Global Positioning System</i>
GSM	<i>Global System for Mobile</i>
GTN	<i>Graph Transformer Network</i>
IBGE	Instituto Brasileiro de Geografia e Estatística
KNN	<i>k-Nearest Neighbors</i>
LADA	<i>Livestock Activity Detection Algorithm</i>
LDA	<i>Linear Discriminant Analysis</i>
MAPA	Ministério da Agricultura, Pecuária e Abastecimento
MLP	<i>Multilayer Perceptron</i>
PCA	<i>Principal Component Analysis</i>
ReLU	<i>Rectified Linear Unit</i>

RF	<i>Random Forest</i>
RNN	<i>Recurrent Neural Network</i>
SMOTE	<i>Synthetic Minority Oversampling Technique</i>
SVM	<i>Support Vector Machine</i>
SOM	<i>Self-Organizing Map</i>
TDF	Transformada Discreta de Fourier
t-SNE	<i>t-Distributed Stochastic Neighbor Embedding</i>
UFMS	Universidade Federal de Mato Grosso do Sul
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo
XGBoost	<i>Extreme Gradient Boosting</i>

Sumário

1	INTRODUÇÃO	16
1.1	MOTIVAÇÃO	18
1.2	OBJETIVOS	19
1.3	ORGANIZAÇÃO DO TRABALHO	20
2	FUNDAMENTAÇÃO TEÓRICA	21
2.1	COMPORTAMENTO ANIMAL	21
2.2	SENSORES	23
2.3	APRENDIZAGEM DE MÁQUINA	24
2.3.1	Pré-processamento	24
2.3.1.1	Normalização	25
2.3.1.2	Técnicas de reamostragem	25
2.3.2	Extração de características	27
2.3.3	Classificação	27
2.3.4	Medidas de avaliação	30
2.3.5	Overfitting	31
2.4	APRENDIZAGEM PROFUNDA	32
3	TRABALHOS RELACIONADOS	37
3.1	CONSTRUÇÃO DAS BASES DE DADOS	37
3.2	PROCESSAMENTO DOS DADOS	40
4	MATERIAIS E MÉTODO	47
4.1	BASE DE DADOS	47
4.2	MÉTODO	52
5	ARQUITETURAS	55
5.1	EXPANSÃO DAS CARACTERÍSTICAS POR COMBINAÇÃO	56
5.2	EXPANSÃO DAS CARACTERÍSTICAS PELA MATRIZ DE ATIVAÇÃO	58
5.3	OTIMIZAÇÃO	60
6	RESULTADOS E DISCUSSÕES	65
6.1	COMBINAÇÃO x MATRIZ DE ATIVAÇÃO X CONV1D	65
6.2	CNNs x CLASSIFICADORES DE APRENDIZAGEM DE MÁQUINA	68
6.3	DISCUSSÕES	72

7 CONCLUSÃO	74
7.1 CONTRIBUIÇÕES	75
7.2 TRABALHOS FUTUROS	76
REFERÊNCIAS	77
GLOSSÁRIO	83
APÊNDICES	84
APÊNDICE A Arquiteturas das CNNs após otimização	85
APÊNDICE B Matrizes de confusão da CNN com abordagem por combinação	86
APÊNDICE C Matrizes de confusão da CNN com abordagem por matriz de ativação	87

1 INTRODUÇÃO

O agronegócio é um dos pilares da economia brasileira. Segundo o Instituto Brasileiro de Geografia e Estatística (IBGE), o setor agropecuário apresentou crescimento de 0,6% no primeiro trimestre de 2020 em comparação ao quarto trimestre de 2019, sendo o único da atividade econômica nacional a crescer nesse período (AGRICULTURA, 2021). Conforme mostra a Figura 1, só a exportação brasileira do setor de carnes movimentou mais de US\$ 19 bilhões, sendo a China o maior comprador do produto brasileiro. Parte dessa iminência se deve aos esforços da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), empresa vinculada ao Ministério da Agricultura, Pecuária e Abastecimento (MAPA), a fim de atingir a transformação da agropecuária com foco na geração de conhecimento e tecnologia, tornando o Brasil um dos maiores produtores e exportadores mundiais de alimentos básicos.

No cenário mundial, até 2050, o aumento previsto de 70% do consumo de produtos de origem animal, tais como carne e leite, agravará ainda mais o problema da fome (BERCKMANS, 2017). O crescimento populacional e a globalização de culturas que incentivam o consumo desses alimentos são alguns dos fatores que contribuem para tal problema.

A fim de remediar essa demanda, conseqüentemente, haverá um aumento da quantidade de animais nas fazendas que acarretará numa maior complexidade do manejo, uma vez que propostas que envolvam uma reeducação alimentar tendem a ser soluções mais complexas nessas culturas. Em virtude disso, surge a seguinte questão: quais estratégias podem promover o manejo animal eficiente e sustentável a fim de mitigar essa situação?

Em especial, a pecuária de precisão, ciência multidisciplinar que envolve a colaboração de áreas como medicina veterinária, fisiologia animal, etologia e engenharias para realizar o manejo da produção ambiental (BERCKMANS, 2017), tem promovido mudanças significativas nos processos de produção e na redução dos impactos ambientais (LU; YOUNG, 2020). O principal objetivo dessa abordagem é maximizar a alocação de recursos específicos e individuais com base em informações precisas para reduzir a perda de diversidade natural gerada pela má gestão dos recursos (BECCIOLINI et al., 2018; SMITH et al., 2015a). BERNARDI et al. (2017) citam as seguintes vantagens da implementação da pecuária de precisão: melhoria e identificação dos níveis de bem-estar animal, redução

de emissão de gases de efeito estufa, facilidade na segmentação dos produtos de origem animal, redução do comércio ilegal e a otimização do retorno econômico das atividades pecuárias.

Nessa perspectiva, tal ênfase no animal empregada pela pecuária de precisão se difere dos métodos tradicionais baseados em informações de rebanho (RAHMAN et al., 2018; RAHMAN et al., 2016; DECANDIA et al., 2017; SMITH et al., 2015a). Por exemplo, a observação visual humana *in loco* do comportamento de bovinos ainda é uma prática comum usada para identificar indivíduos que precisam de uma avaliação veterinária (RODRIGUEZ-BAENA et al., 2020). Essa prática, além de causar estresse aos animais devido à presença humana, é bastante custosa para o criador e, quando aplicada a grandes ambientes ou locais com um grande número de animais, esse e outros métodos tradicionais acabam se tornando difíceis e ineficientes (BENAISSA et al., 2019a; RODRIGUEZ-BAENA et al., 2020). Portanto, o suporte para a tomada de decisão no manejo do rebanho, por meio de informações precisas, torna a pecuária de precisão cada vez mais relevante no planejamento tático e estratégico dos negócios (WILLIAMS et al., 2019b).

Nota-se que os desafios para a consolidação da pecuária de precisão não se limitam apenas em aspectos de investimento financeiro, mas também em uma mudança na cultura presente, principalmente, em pequenas e médias propriedades que ainda se baseiam em métodos tradicionais (PHI et al., 2019). BERNARDI et al. (2017) citam a necessidade de um maior envolvimento das partes interessadas, como instituições de pesquisa, indústria, bem como, a criação de políticas públicas e de incentivo como alternativas para transformação desse cenário.

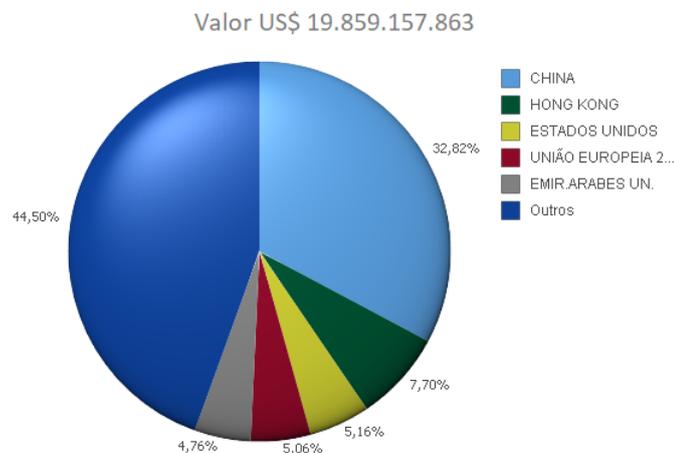


Figura 1 – Exportações Brasileiras do setor de carnes em 2021

Fonte: AGROSTAT¹

1.1 MOTIVAÇÃO

Para a viabilidade dessa abordagem, geralmente, faz-se uso de sistemas de monitoramento contínuo, os quais compreendem desde a captura de dados brutos até a identificação de comportamentos por meio de algoritmos de aprendizagem de máquina. Essas informações precisas e individuais possibilitaram aos produtores obter uma melhor compreensão dos aspectos relacionados às necessidades individuais dos animais, tais como bem-estar (DUTTA et al., 2015; RODRIGUEZ-BAENA et al., 2020; PHI et al., 2019) e desempenho reprodutivo (BECCIOLINI et al., 2018; DECANDIA et al., 2017). Além disso, novas oportunidades de estudos surgiram nas áreas de fenotipagem animal (SMITH et al., 2015b) e seleção genética (RAHMAN et al., 2018).

As tecnologias digitais, por sua vez, estão presentes na maioria desses sistemas, pois permitem uma tomada de decisão mais ágil por meio do monitoramento em tempo real da interação do animal com o meio ambiente e também em conjunto com os outros animais (RAHMAN et al., 2016; PHI et al., 2019; KASFI et al., 2016). Dessa forma, problemas que requerem certa urgência, como a identificação precoce de doenças, podem ser mitigados, sobretudo, em locais com grande quantidade de animais, onde existe um maior risco de perdas.

Em geral, essas soluções consistem em um conjunto de um ou mais sensores que são colocados diretamente sobre o corpo do animal para capturar, armazenar e enviar dados para processamento *a posteriori* (KASFI et al., 2016; DUTTA et al., 2015). Essa adoção das tecnologias digitais, nesse âmbito, também é impulsionada pela competição comercial que cada vez mais exige dos produtores a otimização dos processos operacionais para o aumento da produção (HAMILTON et al., 2019).

Algoritmos de aprendizagem de máquina são amplamente usados para classificar o comportamento de bovinos (RODRIGUEZ-BAENA et al., 2020; WILLIAMS et al., 2019b). Em geral, as propostas se concentram no uso de algoritmos bem conhecidos na literatura, tais como *Support Vector Machine* (SVM - Máquinas de Vetor de Suporte) e *Linear Discriminant Analysis* (LDA), e há pouca exploração da aprendizagem profunda, apesar dela já ter sido aplicada com sucesso em outros problemas no domínio agrícola, como o reconhecimento de plantas, a contagem de frutas e a análise da cobertura do

¹ Estatísticas de Comercio Exterior do Agronegócio Brasileiro - <<https://indicadores.agricultura.gov.br/agrostat/index.htm>>

solo (KAMILARIS; PRENAFETA-BOLDÚ, 2018). Em especial, o sucesso da CNN em diferentes contextos motivou Kasfi et al. (2016) a usar essa técnica para a extração automática de características a fim de discriminar comportamentos a partir de dados de um acelerômetro embarcado em uma coleira. No entanto, na prática, o alto custo computacional e infraestrutura necessária acabam se tornando gargalos para a implantação prática dessa abordagem.

Perante o exposto, esta dissertação se situa no contexto da aprendizagem profunda visando a classificação do comportamento de bovinos por meio da CNN. Além da análise das performances de CNNs conhecidas na literatura, tais como a Lenet (LECUN et al., 2015), a AlexNet (ALOM et al., 2018), a MiniVGG (ISMAIL et al., 2019), a Resnet (HE et al., 2016) e a Mobilenet (HOWARD et al., 2017), foram desenvolvidas outras arquiteturas de CNNs para os diferentes cenários propostos. Dentre essas arquiteturas propostas, também é analisada a técnica de convolução CONV1D sobre o vetor de características originário, além da própria investigação da CONV2D sobre as abordagens de expansão do vetor de características.

1.2 OBJETIVOS

O principal objetivo deste trabalho é explorar o uso da abordagem da aprendizagem profunda, por meio da CNN, para a classificação de seis classes de comportamentos (pastando, andando, em pé parado, em pé ruminando, deitado parado, deitado ruminando) de bovinos da raça Caracu a partir de dados de sensores inerciais embarcados em uma coleira.

Para atingir esse objetivo geral, foram definidos os seguintes objetivos específicos:

- Analisar a estrutura dos dados de cada sensor associado a cada classe de comportamento;
- Avaliar o impacto de técnicas de reamostragem no desempenho dos modelos devido ao desbalanceamento da base de dados;
- Propor duas abordagens para expansão a partir do vetor de característica originário por combinação e pela matriz de ativações;
- Analisar e comparar os desempenhos entre as abordagens das CNNs e entre os classificadores de aprendizagem de máquina mais utilizados na classificação do

comportamento de bovinos;

1.3 ORGANIZAÇÃO DO TRABALHO

O restante deste trabalho está organizado da seguinte forma: o Capítulo 2 aborda os principais conceitos teóricos relacionados ao monitoramento e classificação do comportamento de bovinos; o Capítulo 3 reporta os trabalhos que foram utilizados como base para o desenvolvimento desta dissertação; o Capítulo 4 apresenta as ferramentas utilizadas, a construção e estrutura da base de dados junto com a heurística definida nos experimentos; no Capítulo 5, são detalhadas as arquiteturas otimizadas das CNNs baseadas nas duas abordagens propostas para a expansão do vetor de características originário; o Capítulo 6 expõe os resultados obtidos pelos modelos junto com as análises dos mesmos; por fim, no Capítulo 7, são relatadas as conclusões e listados os trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os fundamentos teóricos relevantes para a construção de uma visão holística sobre o contexto do monitoramento e classificação do comportamento de bovinos. A Seção 2.1 relata a importância do monitoramento de bovinos bem como a descrição das principais classes de comportamento. A Seção 2.2 descreve sobre os sensores e suas aplicações na captura dos dados para representação dos comportamentos. A Seção 2.3 apresenta as principais etapas de um sistema de aprendizagem de máquina, assim como, as técnicas de pré-processamento junto com a definição das medidas de avaliação usadas neste trabalho. Por fim, a Seção 2.4 expõe alguns conceitos básicos sobre a aprendizagem profunda e a CNN.

2.1 COMPORTAMENTO ANIMAL

A saúde e o bem-estar animal contribuem significativamente para a qualidade do produto animal. Uma das justificativas da importância desses elementos pode ser respaldada pela criação do regulamento da União Europeia 2016/429 (*Animal Health Law*), o qual entrou em vigor a partir de 21 de abril de 2021. O bem-estar do animal, em especial, pode ser avaliado pelo monitoramento de diferentes variáveis, tais como a taxa de ventilação, o provimento de ração, o aquecimento e resfriamento do ambiente. Entretanto, segundo Berckmans (2014), poucas ferramentas focam no ator mais importante desse contexto: o animal, visto que a mensuração dos comportamentos habituais também pode retratar o bem-estar dos animais. Por exemplo, as frequências dos comportamentos “pastar”, “ruminar” e “ócio” são alguns dos principais indicadores do desempenho dos bovinos (RIABOFF et al., 2020).

Além disso, esse monitoramento dos comportamentos de bovinos também pode ser usado como uma estratégia para detectar doenças precocemente, posto que mudanças relacionadas às atividades exploratórias, alimentares e reprodutivas são indícios de quando os animais estão doentes (DUTTA et al., 2015; BENAÏSSA et al., 2019a; RAHMAN et al., 2016; RIABOFF et al., 2020).

A claudicação, por exemplo, é uma doença comum e sua detecção precoce reduz significativamente as perdas econômicas relacionadas à produção de leite, os custos do

tratamento veterinário e até mesmo a redução do risco de abate do próprio animal (BENAÏSSA et al., 2019a). Nesse caso, a frequência do comportamento “ócio” é uma variável interessante a ser considerada na detecção da claudicação, pois vacas com essa patologia passam mais tempo deitadas devido ao desconforto causado pela doença (SMITH et al., 2015a; YUNTA et al., 2012; NORRING et al., 2014).

A mastite é outra doença presente e crítica na indústria de laticínios, pois algumas das suas consequências econômicas abarcam a redução da produção e qualidade do leite junto com o aumento do custo por medicamentos. Como agravante, os produtores de leite nem sempre estão cientes das melhores práticas para o seu controle, pois, além da infecção bacteriana, existem outros fatores de risco associados à mastite tais como o ambiente e o manejo. Logo, um manejo adequado pode reduzir a níveis satisfatórios a ocorrência dessa doença (RAHMAN et al., 2009).

A ruminação, processo que consiste em movimentos dinâmicos de mastigação e regurgitação para auxiliar na digestão e absorção de nutrientes, é um dos comportamentos mais importantes levados em consideração na identificação precoce de doenças na pecuária, pois animais doentes comem menos e, conseqüentemente, acabam reduzindo a frequência da ruminação durante tal período (BENAÏSSA et al., 2019a; RAHMAN et al., 2018; HAMILTON et al., 2019). Ademais, esse mapeamento dos comportamentos relacionados à ingestão pode também contribuir para gestão do produtor acerca da quantidade de ração, pasto e suplementos necessários a fim de otimizar o consumo e saciedade dos bovinos (RAHMAN et al., 2018).

Beber é outro comportamento biologicamente importante, pois a periodicidade e a quantidade do consumo diário de água afeta a ingestão de ração, ganho de peso e produção de leite (WILLIAMS et al., 2019a). Em geral, neste comportamento, o animal assume uma posição com a cabeça baixa e uma postura relativamente estática para poder ingerir a água pela boca (PHILLIPS, 2008). Não obstante, tal comportamento ainda não é recorrente entre as soluções de monitoramento de animais (WILLIAMS et al., 2017; WILLIAMS et al., 2019a).

Em outra perspectiva, o monitoramento do comportamento também é uma forma útil de detectar eventos reprodutivos (RIABOFF et al., 2019b). Dado que vacas apresentam maior nível de agitação durante o período de cio (RAHMAN et al., 2018; RAHMAN et al., 2016), o aumento da frequência de comportamentos que mapeiam essa característica, como

“caminhar” ou “correr”, torna-se uma boa escolha para alcançar eficiência na detecção do cio. Como resultado disso, o produtor consegue otimizar a taxa de reprodução por meio da inseminação artificial mais precisa e, conseqüentemente, obter um aumento na produção de leite (BENAISSA et al., 2019a).

2.2 SENSORES

Sensores são dispositivos que capturam os dados que são usados para representar o comportamento dos animais. Idealmente, esses dados devem permitir a discriminação de cada comportamento do animal mediante a captura precisa e individualizada das amostras associadas a cada uma das classes a serem monitoradas (RIABOFF et al., 2020). Além disso, esses sensores otimizam o tempo de trabalho reduzindo a necessidade de observação humana sobre outras variáveis, tais como temperatura, umidade e a localização dos animais.

O uso de acessórios vestíveis como coleiras, brincos, pedômetros e cabrestos colocados no próprio corpo do animal é uma prática comum na maioria das soluções. Sensores são embarcados nesses acessórios a fim de capturar dados brutos sobre o comportamento animal em ambientes onde a visualização *in loco* não é possível ou difícil para os produtores (DECANDIA et al., 2017). O crescimento das vendas de coleiras e pedômetros para a detecção do cio é um indicativo da aceitação dessas tecnologias nesse contexto (HAMILTON et al., 2019; RESEARCH; MARKETS, 2017). O mercado de colares inteligentes destinados ao monitoramento de bovinos foi avaliado em US\$ 243,1 milhões em 2019 e deve crescer para US\$ 513,8 milhões em 2027 (LINKER, 2021).

À vista disso, soluções automatizadas para classificação do comportamento de bovinos a partir de dados de acelerômetros têm aumentado ao longo dos anos devido à obtenção de acurácia satisfatórias (RIABOFF et al., 2019b; RIABOFF et al., 2020). Em especial, dados de acelerômetros embarcados em cabrestos permitem uma boa representação dos comportamentos relacionados à ingestão, pois esses comportamentos apresentam um alto nível de agitação causada pelos movimentos da cabeça e mastigação (BENAISSA et al., 2019a; DECANDIA et al., 2017). Além disso, os acelerômetros também têm sido eficazes na discriminação de comportamentos como “deitar”, “em pé” e “andar” (RIABOFF et al., 2020; WILLIAMS et al., 2019a). Até o final de 2027, o mercado global de acelerômetros

deve atingir US\$ 2.058,7 milhões (WORLD, 2021).

Outros sensores também são usados nos sistemas de monitoramento de animais, como *Global Positioning System* (GPS) (BATISTA et al., 2019), magnetômetros (BISHOP-HURLEY et al., 2014), giroscópios (ANDRIAMANDROSO et al., 2017) e microfones (UNGAR; RUTTER, 2006) que captam dados com representatividade e domínios diferentes (SMITH et al., 2015a; BECCIOLINI et al., 2018).

2.3 APRENDIZAGEM DE MÁQUINA

As soluções atuais reúnem dados provindos de sensores e aplicam técnicas de aprendizagem de máquina supervisionada para classificar comportamentos biologicamente importantes (WILLIAMS et al., 2019b). A aprendizagem supervisionada é uma abordagem em que o algoritmo, durante a fase de treinamento, tem acesso à informação da classe atribuída a cada padrão (BISHOP, 2006). No contexto do comportamento animal, esses dados brutos dos sensores, em um determinado momento no tempo, são associados a uma classe, como por exemplo “comer”, “ruminar” ou “beber”. A Figura 2 mostra um fluxograma clássico de um sistema de aprendizagem de máquina para classificação. Cada uma dessas etapas é brevemente descrita a seguir.

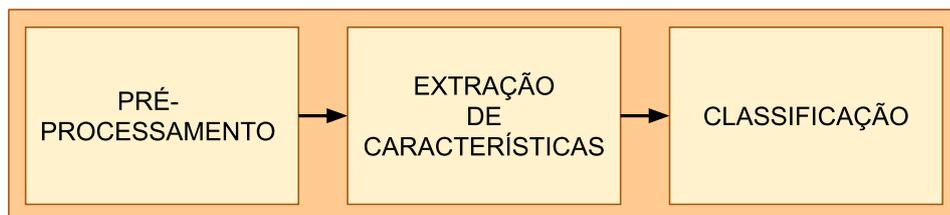


Figura 2 – Fluxograma de um sistema de classificação tradicional de aprendizagem de máquina

2.3.1 Pré-processamento

A etapa de pré-processamento é essencial para contornar situações em que há presença de ruído ou perda de dados, as quais podem ser ocasionadas por diversos fatores durante a captura dos dados brutos pelos sensores. Seu objetivo geral é aplicar transformações nos dados a fim de possibilitar uma extração de características mais eficiente.

Além disso, uma prática comumente utilizada é a segmentação dos dados. No contexto da classificação do comportamento de bovinos, essa segmentação consiste em dividir os dados em estruturas conhecidas como janelas dado um intervalo de tempo. Após essa segmentação, ocorre a extração de características para cada uma dessas janelas.

Essas transformações também podem ser aplicadas para reduzir o custo computacional do modelo. Por exemplo, se os dados de entrada possuem alta dimensionalidade, conseqüentemente, serão necessários mais recursos computacionais e, em alguns casos, até mesmo se torna inviável o processamento desses dados. Diante disso, técnicas de redução de dimensionalidade, como *Principal Component Analysis* (PCA) (MAĆKIEWICZ; RATAJCZAK, 1993), *autoencoders* (BANK et al., 2020) e *t-Distributed Stochastic Neighbor Embedding* (t-SNE) (MAATEN; HINTON, 2008) podem ser opções interessantes. Por outro lado, deve-se ter cuidado para não descartar informações que podem ser importantes para a discriminação de classe (BISHOP, 2006). Além disso, outras técnicas são propostas com diferentes propósitos nesta etapa. Nas subseções a seguir, destacamos algumas utilizadas neste trabalho.

2.3.1.1 Normalização

As características podem possuir domínios e escalas diferentes e dessa forma acabam provocando algumas inconstâncias ou lentidão na convergência da generalização durante o processo de treinamento, pois determinados algoritmos pressupõem que as características tenham variabilidade dentro de uma mesma escala. Diante disso, o uso da normalização dos dados, representada pela Equação 2.1, é uma opção a ser adotada.

$$X = \frac{X - \bar{X}}{s} \quad (2.1)$$

onde X é o dado original, \bar{X} a média e s o desvio-padrão.

2.3.1.2 Técnicas de reamostragem

Outro problema recorrente nas bases de dados é o desbalanceamento dos dados, o qual ocorre quando a quantidade de amostras de uma determinada classe é

significativamente maior que as demais e, portanto, estimula a presença de viés no modelo prejudicando sua generalização.

A fim de mitigar esse problema, surgem técnicas que abordam essa questão basicamente por duas maneiras: ou atribuir custos distintos às amostras das diferentes classes ou realizar uma reamostragem nos dados, seja por subamostragem da classe majoritária, seja por superamostragem da classe minoritária (CHAWLA et al., 2002). Dentre as diversas técnicas de reamostragem, tais como *Adaptive Synthetic Sampling* (ADASYN) (HE et al., 2008), *Borderline-SMOTE* (HAN et al., 2005), *SVM-SMOTE* (TANG et al., 2008), nesta dissertação, foram utilizadas as duas seguintes:

- *Bootstrap*: esse método, definido formalmente por (EFRON; TIBSHIRANI, 1994), consiste em realizar a reamostragem, de maneira aleatória e com reposição, baseado nas amostras originais presentes na base. Além da sua simplicidade, por não requerer um entendimento muito detalhado sobre a distribuição dos dados, essa técnica apresenta bastante potencial para o suporte de diversos propósitos estatísticos como, por exemplo, o cálculo de um intervalo de confiança (GRUNKEMEIER; WU, 2004).
- *Synthetic Minority Oversampling Technique* (SMOTE): método proposto por (CHAWLA et al., 2002) e, diferentemente do *bootstrap*, implementa a reamostragem pela criação de amostras sintéticas. Segundo os autores, a reamostragem com repetição não apresenta melhorias significativas para o desempenho em relação às classes minoritárias e, com respeito às regiões de fronteiras de decisão dos classificadores para as classes minoritárias, a tendência com a adição de dados repetidos é acentuar essas regiões fazendo com que o classificador tenha uma maior predisposição ao *overfitting*.

Essa geração de dados sintéticos opera diretamente no espaço originário de características, ou seja, nenhuma transformação nos dados é necessária para a técnica operar. O Algoritmo 1 ilustra as etapas executadas pelo SMOTE para a criação das amostras sintéticas. Após escolher uma amostra da classe minoritária (linha 12), são computados os seus k vizinhos mais próximos da classe minoritária por meio de uma medida de similaridade dos dados, a qual por padrão é a distância euclidiana (linha 13). Apesar dos autores empregarem o valor de k igual a 5, k é um dos hiperparâmetros do algoritmo e pode ser ajustado por meio de uma heurística a fim de obter melhores resultados. Após isso, é escolhido, aleatoriamente, apenas um

desses vizinhos (linha 18) e gerada uma nova amostra por meio de uma combinação convexa entre a amostra inicialmente escolhida e esse vizinho (linhas 19 a 22).

2.3.2 Extração de características

Durante essa etapa, o objetivo é sintetizar os dados por meio do cálculo de determinadas características a fim de obter uma representabilidade dos padrões com as suas respectivas classes (DUDA et al., 2001). Bishop (2006) define esse estágio como a aplicação de transformações a fim de obter um novo espaço de características que maximize a variabilidade entre as classes distintas. A escolha dessas características pode ser feita manualmente, com base na experiência do especialista, ou por um método automatizado. A utilização de um método para a seleção das características, além de melhorar o desempenho dos modelos, ajuda também a entender a estrutura das particularidades das diferentes classes de comportamento a serem discriminadas (SMITH et al., 2015b). A saída desta etapa serve como entrada para o treinamento dos algoritmos de classificação.

2.3.3 Classificação

Na última etapa, ocorre propriamente a classificação dos padrões, o que consiste em atribuir um rótulo a um determinado padrão, entre um número finito de rótulos discretos, com base nas relações entre as características obtidas na etapa anterior (BISHOP, 2006). Diferentes classificadores já foram abordados para a identificação do comportamento de bovinos, como SVM (KASFI et al., 2016; RAHMAN et al., 2016; HAMILTON et al., 2019), LDA (DECANDIA et al., 2017; BECCIOLINI et al., 2018; RIABOFF et al., 2020), Decision Tree (DT - Árvore de Decisão) (SMITH et al., 2015b; WILLIAMS et al., 2019a; BENAÏSSA et al., 2019a; PHI et al., 2019; RIABOFF et al., 2019b), *Naive Bayes* (NB) (DUTTA et al., 2015; SMITH et al., 2015a; WILLIAMS et al., 2019b) e *Random Forest* (RF) (RAHMAN et al., 2018; RIABOFF et al., 2019a; RIABOFF et al., 2020).

Após uma pesquisa na literatura sobre a classificação do comportamento de bovinos, identificamos e selecionamos os cinco classificadores de aprendizagem de máquina mais utilizados. O resultado dessa lista é bastante compreensível, pois contém classificadores conhecidos, tais como NB, DT, LDA e outros versáteis para diversos contextos tais como

Algoritmo 1: SMOTE(T, N, k)

Entrada: Número de amostras de classe minoritária T ; Porcentagem ($N\%$) para reamostragem SMOTE; Número de k vizinhos mais próximos

Saída: $(N/100) * T$ amostras de classe minoritária sintética

```

1 início
   // (Se  $N$  for menor que 100%, randomize as amostras da classe
   // minoritária, pois apenas uma porcentagem aleatória delas será
   // reamostrada.)
2 se  $N < 100$  então
   // Randomize as  $T$  amostras da classe minoritária
3    $T \leftarrow (N/100) * T$ 
4    $N \leftarrow 100$ 
5 fim
   // O valor de  $N$  é definido em múltiplos inteiros de 100.
6    $N \leftarrow (int)(N/100)$ 
7    $k$ : Número de vizinhos mais próximos
8    $numattrs$ : Número de atributos
9    $Sample[ ]$ : Array para as amostras originais de classe minoritária
10   $newindex$ : Contador inicializado em 0 para o número de amostras sintéticas
   geradas
11   $Synthetic[ ]$ : Array para as amostras sintéticas
   // Calcule os  $k$  vizinhos mais próximos para cada amostra da
   // classe minoritária.
12 para  $i \leftarrow 1$  até  $T$  faça
13   Calcule  $k$  vizinhos mais próximos para amostra  $i$  e salve os índices na
   variável  $nnarray$ 
14    $Populate(N, i, nnarray)$ 
15 fim
   // Função para gerar as amostras sintéticas.
16 Função  $Populate(N, i, nnarray)$ :
17   enquanto  $N \neq 0$  faça
18     Escolha um número aleatório entre 1 e  $k$  e o atribua à variável  $nn$ . Esta
     etapa escolhe um dos  $k$  vizinhos mais próximos da amostra  $i$ .
19     para  $attr \leftarrow 1$  até  $numattrs$  faça
20        $dif \leftarrow Sample[nnarray[nn]][attr] - Sample[i][attr]$ 
21        $gap \leftarrow$  Número randômico entre 0 e 1
22        $Synthetic[newindex][attr] \leftarrow Sample[i][attr] + gap * dif$ 
23     fim
24      $newindex \leftarrow newindex + 1$ 
25      $N \leftarrow N - 1$ 
26   fim
27 fim

```

RF e SVM. A seguir, uma descrição sobre cada um desses classificadores:

- NB: se baseia na independência condicional entre as características. Tal suposição raramente é verdadeira para problemas do mundo real, mas pode ser uma opção interessante se essas dependências forem uniformemente distribuídas em relação às classes ou se cancelarem umas com as outras (ZHANG, 2004). Apesar disso, o NB consegue competir bem com outros classificadores mais sofisticados (RISH et al., 2001).
- DT: método não paramétrico que se baseia em relações entre as características para criar sua estrutura de árvore (raíz e folhas). Possui a vantagem de facilitar a interpretação das classificações por meio da visualização da árvore. Mesmo sendo um método que não requer muito recurso computacional, as DTs são capazes de resolver problemas difíceis e práticos (QUINLAN, 1986).
- LDA: além da classificação, é utilizada para redução de problemas com alta dimensionalidade e visualização de dados. Esse método consiste em determinar um conjunto de vetores de projeção por meio da maximização da matriz de dispersão interclasse e minimizando a matriz de dispersão intraclasse no espaço de características (WANG; TANG, 2004).
- RF: esse método consiste em uma combinação de DTs, em que cada uma é gerada por meio de uma seleção aleatória de características e amostras. Cada DT classifica a amostra em uma determinada classe e, no fim, a RF retorna o rótulo que obteve a maior quantidade de classificações dentre esse conjunto de DTs (PAL, 2005).
- SVM: esse método mapeia as entradas em um espaço de características de maior dimensão por meio de uma transformação não linear, no qual é construído um hiperplano com suas margens maximizadas e com a fronteira de decisão linear provendo uma alta capacidade de generalização (CORTES; VAPNIK, 1995).

Além disso, o uso de abordagens que levam em consideração a combinação de classificadores ganhou destaque em diferentes contextos. Essas propostas usam vários modelos a fim de melhorar o desempenho que seria obtido se fosse utilizado apenas um modelo individual (DUTTA et al., 2015).

2.3.4 Medidas de avaliação

A utilização de medidas de avaliação é fundamental no desenvolvimento de um sistema de aprendizagem de máquina, pois permite analisar a qualidade da generalização dos modelos e, conseqüentemente, direcionar as decisões do projetista na definição dos valores dos hiperparâmetros e concepção do modelo. Apesar de haver diversas medidas disponíveis, a escolha da medida a ser empregada depende da natureza do problema junto com as especificidades dos dados. Logo, o sucesso do projeto, na prática, pode ser comprometido caso o especialista tenha escolhido medidas equivocadas durante o seu desenvolvimento.

Devido a sua simplicidade e clareza na representação da performance de um modelo, a matriz de confusão é uma ferramenta bastante utilizada. A Figura 3 ilustra a estrutura de uma matriz de confusão. Nota-se que são mapeadas todas as previsões do modelo, onde a diagonal principal, em particular, representa todos os acertos; os erros, os demais valores. Basicamente, a matriz indica quantas amostras pertencem a cada um dos seguintes grupos:

- Verdadeiro Positivo (VP): amostras positivas que o modelo classificou corretamente como positivas;
- Falso Positivo (FP): amostras negativas que o modelo classificou erroneamente como positivas;
- Verdadeiro Negativo (VN): amostras negativas que o modelo classificou corretamente como negativas.
- Falso Negativo (FN): amostras positivas que o modelo classificou erroneamente como negativas;

		VALOR PREDITO	
		SIM	NÃO
VALOR REAL	SIM	VP	FN
	NÃO	FP	VN

Figura 3 – Estrutura da matriz de confusão

Por apresentar uma grande quantidade de informações, algumas métricas podem ser extraídas a partir de uma matriz de confusão. Nesta dissertação, destacamos as seguintes

métricas derivadas:

- Acurácia: porcentagem de todas as amostras previstas corretamente pelo modelo.

$$\frac{VP + VN}{VP + VN + FP + FN} \quad (2.2)$$

- Precisão: porcentagem das amostras classificadas corretamente como positivas em relação a todas as amostras classificadas como positivas.

$$\frac{VP}{VP + FP} \quad (2.3)$$

- Sensibilidade: porcentagem das amostras classificadas corretamente como positivas em relação a todas as amostras positivas.

$$\frac{VP}{VP + FN} \quad (2.4)$$

- F1-score: média harmônica entre a precisão e a sensibilidade.

$$2 * \frac{\textit{Precisão} * \textit{Sensibilidade}}{\textit{Precisão} + \textit{Sensibilidade}} \quad (2.5)$$

2.3.5 Overfitting

Um fenômeno bastante comum durante o desenvolvimento de um sistema de aprendizagem de máquina é o *overfitting*, o qual consiste no maior ajuste do modelo apenas em determinados dados acarretando em uma baixa capacidade de generalização para novos dados. A complexidade do algoritmo, ruídos nos dados ou até mesmo o desbalanceamento da base são alguns fatores que contribuem para sua origem. Portanto, tal problema precisa ser tratado, uma vez que compromete a análise das medidas de avaliação as quais não serão correspondidas na prática.

Assim sendo, algumas providências podem ser tomadas para mitigação dessa situação. Uma boa prática é dividir a base de dados em duas partes: treino e validação, de sorte que a análise das medidas de avaliação é focada na base de validação e nenhum dado presente na base de treinamento deverá pertencer à base de validação. Além disso, algumas técnicas são propostas para a remoção do *overfitting*. Dentre elas, destacamos as duas seguintes bastante utilizadas no contexto das redes neurais:

- Regularização L1: esse método induz as redes a aprenderem pesos menores por meio da adição do termo de regularização o qual aumenta a penalização na atualização

do vetor de pesos. A Equação 2.6 representa o cálculo da função de perda (*loss*) com a adição do termo de regularização L1. Além disso, como a sua derivada resulta em um escalar, esse tipo de regularização penaliza os pesos de forma linear independentemente do vetor de pesos, resultando em uma redução constante e aproximando os pesos pequenos (conexões menos importantes) para zero, ou seja, essa técnica acaba anulando algumas conexões e, conseqüentemente, faz o papel de um seletor de características.

$$\mathcal{J} = Loss(y - \hat{y}) + \lambda \sum_1^n |w_i| \quad (2.6)$$

- Dropout: proposta por Hinton et al. (2012), essa técnica é utilizada para desativar temporal e aleatoriamente uma porcentagem de neurônios presentes nas camadas ocultas de uma rede neural, a qual passa a ser um hiperparâmetro, durante a etapa de treinamento. Intuitivamente, essas desativações de alguns neurônios permitem que as atualizações dos pesos sejam realizadas com base em diferentes redes e, por sua vez, dificultem o *overfitting* da rede como um todo. Segundo os autores, essa técnica evita a detecção das relações complexas das características e faz com que cada neurônio tenda a detectar as características mais relevantes, uma vez que será atualizado em instantes com arquiteturas de redes diferentes.

2.4 APRENDIZAGEM PROFUNDA

A aprendizagem profunda é uma extensão de um algoritmo de aprendizagem de máquina, o *Multilayer Perceptron* (MLP), com a aplicação de diversas funções que permitem uma representação hierárquica das características em vários níveis de abstração por meio da composição de módulos não lineares, também conhecidos como *perceptrons*, de modo que, ao decorrer das camadas das redes, essa abordagem compreende as relações entre os dados com as suas respectivas classes por meio do algoritmo *backpropagation* (SCHMIDHUBER, 2015). LeCun et al. (2015) definem o algoritmo *backpropagation* como uma aplicação prática da regra da cadeia de derivadas por meio do cálculo do gradiente da função objetivo em relação aos pesos da rede.

Outrora, o desenvolvimento de um sistema de reconhecimento de padrões carecia de uma engenharia detalhada para a extração das características que melhor discriminam os padrões, o que, muitas vezes, acabava se limitando pelo conhecimento e experiência do

projetista. Diante disso, uma das vantagens desta abordagem é a extração automática das características diretamente dos dados brutos evitando essa engenharia manual (LECUN et al., 1998).

Outra benesse da aprendizagem profunda é a concepção de representações distribuídas as quais permitem a generalização para novas características aprendidas além daquelas vistas durante o treinamento. Dessa maneira, cada uma dessas representações pode contribuir para a modelagem de outras representações similares até mesmo em outros domínios, introduzindo o conceito de *transfer learning* (ZHUANG et al., 2020). No entanto, vale enfatizar que os modelos de aprendizagem profunda, comumente, são modelos complexos que requerem o uso elevado de recursos computacionais.

As CNNs são um dos destaques da aprendizagem profunda que trouxeram importantes avanços nas áreas de processamento digital de imagens, áudio e texto (LI et al., 2021). O nome é derivado da execução dos filtros de convolução sobre as entradas, uma vez que essas redes são projetadas para processar dados estruturados em forma de matrizes. Sobre esses filtros de convolução, há uma divisão em relação à quantidade de dimensões que o mesmo possui. Os filtros de uma única dimensão, conhecidos como CONV1D, são filtros comumente utilizados para a análise de séries temporais, exatamente por esse tipo de dados possuir, na sua estrutura, uma única dimensão variante ao decorrer do tempo (ZHONG et al., 2019). Já os filtros com duas dimensões, também conhecidos como CONV2D, são filtros utilizados para matrizes com estrutura bidimensional que levam em consideração a relação entre a espacialidade dos valores presentes na matriz (BIRCH et al., 2020).

LeCun et al. (2015) citam quatro conceitos principais em que as CNNs se baseiam para realizar a extração de características dos dados brutos: conexões locais, pesos compartilhados, *pooling* e o uso de múltiplas camadas. As arquiteturas das CNNs são estruturadas em uma série de estágios. O primeiro compõe dois tipos de camadas, as quais são inspiradas diretamente na noção clássica de células simples e complexas na neurociência visual: camada convolucional e camada de *pooling*.

A função da camada de convolução é identificar as características locais na matriz anterior por meio da execução dos filtros. Diante disso, surgem alguns dos hiperparâmetros das CNNs, tais como o número e a dimensão desses filtros de convolução. Inicialmente, esses filtros são instanciados com valores aleatórios e seus pesos são aprendidos pela própria

rede por meio do algoritmo *backpropagation*.

Após a camada de convolução, é comum a aplicação de uma função de ativação para a transformação não linear nos dados. Dentre as várias funções de ativação, mencionamos a função não linear *Rectified Linear Unit* (ReLU), a qual é representada pela Equação 2.7. Geralmente, a função ReLU permite uma convergência mais rápida em redes com muitas camadas (LECUN et al., 2015). Além disso, essa função, por retornar zero para entradas negativas, acaba não habilitando todos os neurônios ao mesmo tempo, o que resulta em uma rede mais esparsa e eficiente.

$$ReLU(x) = \max(0, x) \quad (2.7)$$

Já a camada de *pooling* mescla as características semelhantes para reduzir a variação delas em aspectos de posição e aparência (LECUN et al., 2015). Ela opera sobre a matriz de ativação da camada anterior reduzindo suas dimensões. É muito comum a utilização das operações de máximo (*max pooling*) e média (*average pooling*) nos filtros. Ao contrário dos filtros de convolução, essa camada não introduz muitos hiperparâmetros no modelo e acaba se tornando uma ferramenta útil no controle do *overfitting*.

Em seguida, é comum adicionar uma ou mais camadas *Fully Connected* (FC), a qual consiste em um conjunto de neurônios totalmente conectados semelhante a uma MLP, em especial, a camada de saída contendo neurônios com a função de ativação *softmax*, definida pela Equação 2.8, a qual formata a saída da rede em um vetor em que cada posição representa uma probabilidade do padrão pertencer a uma determinada classe.

$$softmax(x_i) = \frac{\exp(x_i)}{\sum_j \exp(x_j)} \quad (2.8)$$

Após a obtenção do vetor de probabilidades, ocorre o cálculo da função de perda (*loss*) a qual avalia a saída da rede durante o treinamento. Essa função avalia o quão bem a rede está ajustada aos dados, ou em outras palavras, o quão a rede está aprendendo. Assim, as atualizações dos pesos da rede se baseiam nessa função e na direção oposta do cálculo do vetor gradiente (derivada), o qual indicará a direção e magnitude, a fim de minimizar o valor do *loss*. Logo após, o *backpropagation* propagará esses valores para as

outras camadas junto com as suas respectivas derivadas. Dentre algumas dessas funções, destacamos a *Cross Entropy* (CE) representada pela Equação 2.9.

$$CE(y, \hat{y}) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^c [y_{i,j} \ln \hat{y}_{i,j} + (1 - y_{i,j}) \ln (1 - \hat{y}_{i,j})] \quad (2.9)$$

Onde y é o vetor representando a classe real para cada uma das n amostras, \hat{y} o vetor de saída da rede com relação as c classes.

Isto posto, cinco CNNs conhecidas na literatura foram utilizadas para serem analisadas e comparadas com os resultados das CNNs propostas por essa dissertação. A seguir, uma descrição de cada uma delas:

- **Lenet:** proposta por LeCun et al. (1998), essa CNN se baseia em dois tipos de camadas: camadas de convolução e reamostragem. Sua arquitetura é composta por sete camadas: três camadas de convolução, duas de subamostragem e duas camadas totalmente conectadas. Inicialmente, essa arquitetura foi projetada para resolver o problema de reconhecimento de caracteres manuscritos ou impressos, entretanto devido a sua simplicidade e bom desempenho para outros problemas (WEI et al., 2019; LIN et al., 2016; ZHU et al., 2021), ganhou destaque na literatura.
- **Alexnet:** essa rede possui oito camadas: cinco camadas de convolução e três camadas totalmente conectadas. Introduziu a ideia de sobreposição nas camadas de subamostragem a fim de evitar o *overfitting*, além de mostrar o melhor desempenho da função não linear *ReLU* sobre outras funções bastante utilizadas na literatura, como a tangente hiperbólica (KALMAN; KWASNY, 1992) e a sigmoide (HAN; MORAGA, 1995).
- **MiniVGG:** rede proposta por Ismail et al. (2019) que utiliza as técnicas de normalização em lotes e decaimento da taxa de aprendizagem para melhorar a performance do modelo. É uma variante simplificada da rede VGGNet (SIMONYAN; ZISSERMAN, 2014), a qual é utilizada para processar matrizes de alta dimensão. Desta forma, faz uso filtros de convolução pequenos (3x3), o que segundo os autores, permite com que a rede consiga generalizar para diferentes problemas de classificação.
- **Resnet:** também conhecida como *Residual Network*, essa CNN proposta por He et al. (2016) é uma das redes que exploram a aprendizagem muito profunda introduzindo o conceito de blocos residuais para a construção da rede com o objetivo de mitigar o

problema de explosão e desaparecimento do gradiente.

- Mobilenet: rede proposta por Howard et al. (2017) é formada por uma arquitetura simplificada por meio de convoluções separáveis em profundidades, conceito introduzido por Sifre e Mallat (2014). Todas as camadas são seguidas por uma normalização em lotes e a função *ReLU*, com exceção da última camada que possui a função de ativação *softmax*.

3 TRABALHOS RELACIONADOS

Neste capítulo, são abordados alguns tópicos dos trabalhos relacionados ao monitoramento e classificação do comportamento de bovinos que serviram como base para o desenvolvimento dessa pesquisa. Na Seção 3.1, são apresentados os mecanismos para a aquisição dos dados, relatando o processo de rotulação, os acessórios utilizados junto com os sensores embarcados. A Seção 3.2 aborda as heurísticas e técnicas utilizadas pelas propostas em todas as etapas de um sistema de classificação.

3.1 CONSTRUÇÃO DAS BASES DE DADOS

Para o treinamento de algoritmos de aprendizagem supervisionada no contexto do monitoramento de bovinos, bases de dados precisam ser construídas por meio da captura dos dados provenientes de sensores e posterior rotulação das amostras com suas respectivas classes de comportamentos. Em alguns casos, a rotulagem dos dados é realizada manualmente por meio de visualização humana *in loco*. Esse método pode até pode ser eficaz em casos que há um pequeno número de animais ou em pequenas fazendas, mas se torna impraticável em casos com uma grande quantidade de animais ou grandes propriedades, pois requer um maior grau de atenção, disponibilidade e experiência do especialista durante o processo (RIABOFF et al., 2019b; BECCIOLINI et al., 2018; RODRIGUEZ-BAENA et al., 2020). Além disso, a presença humana pode causar desconforto aos animais e até interferir na espontaneidade dos comportamentos. Dutta et al. (2015), Smith et al. (2015a), Kasfi et al. (2016), Rahman et al. (2016), Benaissa et al. (2019a), Becciolini et al. (2018), Williams et al. (2019b), Riaboff et al. (2019b), Riaboff et al. (2019a), Rodriguez-Baena et al. (2020), Riaboff et al. (2020) são alguns exemplos de trabalhos que utilizaram esse método para a construção das bases.

Dutta et al. (2015), Smith et al. (2015a), Lomba (2015) destacam-se pelo uso de tecnologias digitais através de aplicativos móveis a fim de melhorar a precisão da sincronização dos dados durante o processo de rotulação das amostras. Desta forma, nota-se que as tecnologias digitais são ferramentas interessantes, pois podem reduzir a intervenção humana nesse processo (RODRIGUEZ-BAENA et al., 2020). Outro exemplo é a rotulação realizada pela análise, *a posteriori*, de imagens geradas por câmeras de

monitoramento, a qual se mostra uma alternativa menos invasiva e sujeita a erros. Alguns trabalhos, como Smith et al. (2015b), Williams et al. (2019a), Rahman et al. (2018), Decandia et al. (2017), Phi et al. (2019), utilizaram essa estratégia. Devido à necessidade de uma infraestrutura adequada para a captura das imagens, o que não é uma opção viável em alguns casos, esse método acaba possuindo menor adesão.

Nessa mesma perspectiva de redução do contato humano com o animal, a telemetria é um recurso essencial para um sistema de monitoramento automatizado, pois elimina a atividade de retirada periódica do dispositivo acoplado ao animal para carregar os dados capturados. Decandia et al. (2017) usaram o dispositivo BEHARUM, o qual capturava dados a cada 1 segundo e os enviava para um gateway através de um sistema de rede sem fio LoRa (DEVALAL; KARTHIKEYAN, 2018). Por fim, esse gateway se comunicava com um computador via *Global System for Mobile* (GSM) para posterior processamento dos dados.

As soluções empregam sensores embarcados em acessórios vestíveis de diferentes maneiras, sendo a coleira um dos mais utilizados. Essa preferência é compreensível, pois as coleiras são acessórios práticos que podem ser facilmente fixados e retirados do animal, o que é uma tarefa importante e rotineira para coleta dos dados. Além disso, a coleira tem uma boa capacidade de capturar os movimentos dos principais comportamentos dos bovinos sem causar desconforto. Entretanto, após colocar o acessório no corpo, o animal pode demorar um pouco para se acostumar. Assim, uma boa prática adotada para mitigar esse impacto no comportamento do animal é fixar a coleira, ou qualquer outro acessório, por um período de adaptação antes de realizar efetivamente a coleta dos dados. Alguns trabalhos, como Benaissa et al. (2019a), sugerem colocar esses acessórios um dia antes da coleta de dados, deixando o animal livre para interagir com o ambiente.

Ainda em relação aos acessórios para armazenar os sensores e vestir no corpo do animal, Hamilton et al. (2019) se distinguem por utilizar um acelerômetro embarcado em um dispositivo bolus para classificação da ruminação. Os autores consideram a contração do retículo, principal órgão que participa do processo de ruminação, pois é responsável pela contração que leva à regurgitação.

Enquanto boa parte das soluções usam um ou dois acessórios simultâneos por animal, Rahman et al. (2018) se destacam por equipar cada animal com três acessórios: uma coleira, um cabresto e um brinco. Em geral, os dados gerados pelo cabresto foram

classificados com maior precisão do que os dados gerados pela coleira e brinco. Segundo os autores, o uso do cabresto é uma boa opção para o monitoramento, pois, além do movimento da cabeça, é possível captar o movimento de mastigação do animal.

Além das soluções desenvolvidas pelos próprios pesquisadores, há trabalhos que utilizam soluções comerciais. Benaissa et al. (2019a) fixaram, em cada animal, o dispositivo *RumiWatch* (solução comercial desenvolvida para monitorar o comportamento de ruminância, alimentação e outras atividades em vacas leiteiras) e uma coleira equipada com um acelerômetro desenvolvido pelos próprios pesquisadores. Uma análise do desempenho dos modelos foi realizada conforme a frequência de amostragem dos sensores diminuía.

Em relação aos sensores, o uso de acelerômetros possui uma certa preferência devido a sua melhor capacidade de representação e precisão do comportamento de bovinos (WILLIAMS et al., 2019b). Por exemplo, os dispositivos HOBO *data logger* são um dos sensores presentes na literatura para o estudo do comportamento de bovinos (RODRIGUEZ-BAENA et al., 2020).

Somado a isso, a combinação de dois ou mais sensores, em diferentes posições no corpo do animal, com dados de diferentes domínios e representações é uma opção para o aumento do desempenho e do escopo dos sistemas de monitoramento (RIABOFF et al., 2020). Por exemplo, Williams et al. (2019b) informam que os dados de latitude, longitude e velocidade de um GPS por si só não foram suficientes para classificar alguns comportamentos relevantes, como a ruminância. Outras propostas, como o dispositivo Camazotz (JURDAK et al., 2013; SMITH et al., 2015b), combinam um acelerômetro e um sensor de pressão embarcado em um cabresto.

Além disso, a definição da taxa de amostragem dos sensores, medida que representa a quantidade de amostras que um sinal analógico é capturado em um determinado intervalo de tempo, torna-se uma decisão importante durante o desenvolvimento de um sistema de monitoramento contínuo. Rodriguez-Baena et al. (2020) mostram a importância da frequência de captura do sensor para a representação dos comportamentos. A escolha de uma taxa de amostragem deve ser orientada para alcançar uma boa precisão na representação de comportamentos junto com o menor consumo de energia para permitir uma maior vida útil da bateria (RIABOFF et al., 2020). Por exemplo, apesar de possuir dados de localização via GPS, Smith et al. (2015a) optaram por não utilizar os dados

desse sensor, pois o seu foco era uma solução barata e que consumisse menos energia. Hamäläinen et al. (2011) recomendam o valor de 10Hz para acelerômetros, uma vez que vacas são animais calmos.

Por fim, vale enfatizar a construção da base de dados por Becciolini et al. (2018). Ao contrário de outros trabalhos que possuem uma pequena duração nos experimentos, os autores construíram uma base de dados composta por dados de comportamentos durante 8 meses (abril a novembro), compreendendo um total de 11 dias de observação direta com 120 horas de dados rotulados. Assim, os autores conseguiram mapear uma variabilidade do comportamento de bovinos durante diferentes estações climáticas e com animais de raças diferentes.

3.2 PROCESSAMENTO DOS DADOS

Certos problemas nas bases de dados utilizadas pelos trabalhos foram abordados na etapa de pré-processamento. O ruído nos dados, por exemplo, é um problema habitual que pode ser causado pelo mal funcionamento do sensor ou até mesmo pela reorientação do sensor gerada por um deslocamento físico do dispositivo sobre o corpo do animal (DUTTA et al., 2015). Para remover frequências indesejadas de dados associados ao ruído, o uso de filtros passa-baixa, os quais permitem a passagem de sinais de baixa frequência e atenuam a intensidade de sinais de alta frequência, foi recorrente, pois requer menos processamento quando comparado a outros filtros (RIABOFF et al., 2019b; PHI et al., 2019; BENAÏSSA et al., 2019a). Somado a isso, a normalização dos dados, técnica utilizada para abstrair as diferenças entre os valores das características dos padrões a fim de não prejudicar o processo de aprendizagem, também esteve presente (DUTTA et al., 2015; WILLIAMS et al., 2019a).

Um dos fatores que aumentam a complexidade da classificação é a precisão na captura dos dados exclusivos para cada uma das classes de comportamento. Em geral, o método clássico consiste em realizar a segmentação dos dados em intervalos fixos, conhecidos como janelas, para posterior extração de características (WILLIAMS et al., 2019b). No entanto, na prática, os animais podem expressar diferentes comportamentos dentro de uma mesma janela de tempo, mesmo que uma janela curta seja escolhida para evitar esse problema. Sendo assim, esse método resulta em inconsistências durante o

processo de treinamento e, conseqüentemente, no desempenho do modelo.

Para mitigar essa situação, Riaboff et al. (2020) optaram pelo descarte das janelas que continham mais de um comportamento simultaneamente, porém essa exclusão de dados pode levar à perda de informações importantes para a discriminação dos comportamentos, principalmente, nos casos em que a base de dados é pequena. Já Williams et al. (2019b) propõem uma técnica de segmentação variável que separa as séries temporais em segmentos de tamanhos diferentes com base em pontos de mudança estatisticamente significativos, os quais são indicativos de mudanças no comportamento do animal. Shahriar et al. (2016) também utilizaram essa abordagem, mas no contexto de detecção do cio em vacas leiteiras. Ressalta-se que essas técnicas requerem o uso de sensores capazes de capturar essas alterações nos comportamentos. Além disso, variáveis como a espécie do animal, a taxa de amostragem e a localização do sensor no corpo do animal podem influenciar no desempenho desse tipo de segmentação. Outro ponto a se frisar é a capacidade de executar em dados não rotulados para identificar mudanças no comportamento, logo problemas de aprendizagem não supervisionada, os quais não requerem conhecimento prévio dos rótulos das amostras, podem tirar proveito dessa abordagem.

Outros trabalhos, devido ao forte desequilíbrio do número de amostras entre as classes de comportamento, acabam optando por aglutinar comportamentos semelhantes para a classificação. Benaissa et al. (2019a), Benaissa et al. (2019b) consideraram o comportamento “beber” como “comer”, em vez de remover o pequeno número de amostras do comportamento “beber”. Outra prática para remover o viés da classe majoritária na etapa de classificação do modelo é realizar uma reamostragem excluindo os padrões da classe majoritária até aproximar com o número de padrões entre as classes minoritárias (RAHMAN et al., 2016; DUTTA et al., 2015). Por exemplo, Dutta et al. (2015) realizaram uma subamostragem com 785 amostras para cada uma das cinco classes (pastar, ruminar, descansar, caminhar e outro). Contudo, esse método acaba não se tornando uma alternativa interessante, pois a perda dos dados gerados pode afetar significativamente a generalização dos modelos e, até mesmo, se tornar inviável em situações de base de dados com poucas amostras.

Mesmo assim, outros estudos optaram pelo descarte das classes de comportamento com pouca representação no banco de dados. Devido ao forte desequilíbrio das amostras entre as nove classes de comportamento (pastar, procurar, andar, mastigar, ruminar

deitado ou sentado, ruminar em pé, descansar deitado, descansar em pé com a cabeça erguida e outro), Kasfi et al. (2016) optaram por considerar apenas duas classes: “pastar” e “outro” representando as demais classes. Dessa forma, essa escolha pela redução da quantidade de classes resultou também no aumento da performance dos classificadores. Essa prática também foi usada por Rahman et al. (2018) no que diz respeito às classes “caminhar”, “descansar” e “outros”.

A sobreposição (*overlap*) é um método que consiste em definir uma janela levando em consideração uma determinada porcentagem do início e do fim de duas janelas consecutivas, tornando-se uma opção interessante nos casos em que não existam sequências suficientes para as classes de comportamento (RIABOFF et al., 2019b). Riaboff et al. (2020) executaram uma combinação de configurações de janela entre 5 e 10 segundos com sobreposição de janela de 50% e 90%. Para cada combinação, uma DT foi usada para classificar os comportamentos. A configuração que gerou o melhor resultado foi semelhante à obtida por Riaboff et al. (2019b) com janelas de 10 segundos e 90% de sobreposição.

Como dito anteriormente, depois de segmentar os dados brutos em janelas, dá-se início a extração de características para cada uma dessas janelas. Nesta fase, há a aplicação de uma variedade de técnicas que se agrupam na análise empírica, limitada pelo conhecimento e experiência do pesquisador, ou na extração automática de medidas estatísticas.

Smith et al. (2015b) compararam o uso de três grupos de características para classificar as seis classes de comportamento em bovinos (pastar, descansar, andar, ficar em pé, ruminar e coçar). O grupo de características estatísticas (média, desvio padrão, assimetria, curtose, a soma dos valores absolutos) (NISBETT et al., 2020) é amplamente utilizado por outros trabalhos com a abordagem padrão para a extração de características. Os outros dois grupos eram os das características espectrais (centroide espectral, espalhamento espectral, curtose espectral, coeficiente de achatamento, frequência fundamental, coeficiente autorregressivo espectral, coeficiente autorregressivo) (ZHAO; LIU, 2012) e o das características da teoria da informação (entropia, eficiência de codificação) (BURGIN, 2010). Além disso, uma busca gulosa foi realizada para identificar as características que eram mais eficazes na discriminação entre os comportamentos dos animais. Conforme relatado pelos autores, esse método escolhido é uma das limitações do trabalho, visto que é improvável que esse método de otimização convirja para a solução

ótima global. Eles também sugerem a combinação das características desses três grupos para melhorar a discriminação de comportamentos, uma vez que as características estatísticas representam a forma da distribuição de cada intervalo de tempo, as características espectrais representam a forma do espectro de potência e as características do grupo da teoria da informação representa a previsibilidade de cada intervalo de tempo. Esse estudo conclui que as características estatísticas de segunda ordem (desvio padrão e soma dos valores absolutos), as quais representam o nível de intensidade do movimento, foram as características individuais mais discriminantes entre todas utilizadas nos experimentos.

Além desses grupos de características, outros trabalhos, como Rahman et al. (2018), calculam a Transformada Discreta de Fourier (TDF) (WINOGRAD, 1978) com o objetivo de explorar as características no domínio da frequência. A lista de características extraídas inclui a média, desvio padrão, assimetria e curtose. Nessa mesma perspectiva, Riaboff et al. (2020), Riaboff et al. (2019b) também abordam a extração de características nos domínios do tempo e da frequência.

Em relação aos dados de aceleração, podem ser extraídas características que se agrupam em aspectos estáticos e dinâmicos (RIABOFF et al., 2019b; PHI et al., 2019). O componente estático da aceleração representa a posição do animal por meio da orientação do sensor em relação à gravidade, em que a média e a mediana são algumas representantes desse aspecto. Já o componente dinâmico representa diretamente a intensidade dos movimentos capturados. Variância, desvio padrão, mínimo, máximo e quartis são algumas das medidas utilizadas para representar essa propriedade. Por exemplo, Williams et al. (2019a) usaram a média e a variância dos dados do acelerômetro para representar a posição e o nível de atividade dos bovinos respectivamente. Eles também usaram um modelo mistura linear para determinar quais características representam melhor as diferenças de posição e movimento do comportamento “beber” do restante dos outros comportamentos. Os autores perceberam que a distribuição dos dados para o comportamento “beber” foi muito semelhante ao comportamento “em pé”. No entanto, houve uma diferença significativa do comportamento “beber” com o comportamento “andar”.

Dutta et al. (2015) destacam-se por utilizar técnicas de aprendizagem supervisionada para reduzir a dimensionalidade, compreender e visualizar a estrutura e agrupamento das amostras de cada classe de comportamento. As técnicas utilizadas pelos autores foram PCA, *Self-Organizing Map* (SOM) e *Fuzzy C Means* (FCM). Empiricamente, os autores

testaram os valores no intervalo entre 2 a 20 segundos a fim de obter o valor do tamanho da janela que resultasse na maximização da separação entre os clusters. A janela de 5 segundos foi escolhida e partir dela oito características foram extraídas (negentropia, energia, auto-regressiva, média, área sob a curva, desvio padrão, curtose e assimetria).

O sucesso das CNNs em diferentes contextos motivou Kasfi et al. (2016) a usarem essa técnica de aprendizagem profunda para a extração automática de características para a discriminação de padrões de comportamento a partir de dados de um acelerômetro embarcado em uma coleira. Os autores fazem uma comparação com o trabalho de Rahman et al. (2016), os quais, por sua vez, usam um modelo baseado em características extraídas por um *autoencoder* empilhado. Um *autoencoder* empilhado consiste em uma rede neural com várias camadas de *autoencoders* esparsos (RAHMAN et al., 2016). O modelo com CNN apresentou desempenho semelhante ao trabalho com *autoencoders*, mas demonstrou ser mais eficiente em relação ao tempo.

Rahman et al. (2016) comparam a eficácia da extração automatizada de características usando *autoencoders* com a extração manual de características estatísticas para a discriminação dos comportamentos de bovinos. Média, desvio padrão, assimetria, curtose, mínimo e máximo foram as características estatísticas utilizadas no experimento. A motivação para o uso de *autoencoders* foi a má adaptação dos métodos tradicionais de extração de características usados no treinamento de classificadores na discriminação dos comportamentos em diferentes contextos, tais como espaços físicos, raça do animal e dados de diferentes sensores. Assim, com a convergência para extração de características, não há mais a necessidade de remodelagem na etapa de extração de característica para cada um desses casos.

Com relação à tarefa de classificação propriamente dita, Smith et al. (2015a) destacam-se por propor dois classificadores de séries temporais multivariados inspirados no *Bag-of-Words* (BoW) e no *Bag-of-Features* (BoF) para classificar dez classes de comportamento (pastar, procurar, caminhar, mastigar, ruminar deitado, ruminar em pé, descansar deitado, descansar em pé, cabeça abaixada e outros): o *Bag of Class Posteriors* (BOCP) e o *Bag of Class Posteriors with Ordering* (BOCPO). Os autores também exploram uma limitação de outras propostas que envolvem apenas uma janela de tempo fixa para a análise de dados. Os resultados indicam que os modelos BOCP e BOCPO tiveram um melhor desempenho de 43% a 77% em relação à abordagem de janela

fixa. Essas melhorias se devem a melhor representação dos comportamentos gerados pela redução da interferência de mais de um comportamento na mesma janela.

Por outro lado, outros trabalhos abordam um menor número de classes simultaneamente. Rodriguez-Baena et al. (2020) abordam a classificação de apenas duas classes (atividade e inatividade) propondo o *Livestock Activity Detection Algorithm* (LADA) para detectar e quantificar automaticamente períodos de atividade animal. Os resultados indicam que o método proposto pode auxiliar os produtores no manejo animal como uma pré-etapa de um sistema mais completo que envolva a classificação de mais classes de comportamentos.

Williams et al. (2019a) mostraram que é possível identificar o comportamento “beber” de bovinos de corte a partir de dados de acelerômetro. Houve registro das seguintes classes de comportamento: beber, ficar em pé (cabeça erguida), caminhar e ficar em pé (cabeça baixa). Entretanto, os autores não realizaram uma análise envolvendo todos os comportamentos simultaneamente, mas sim, com base em pares de comportamentos (classificação binária). Logo, essa decisão acaba se tornando uma limitação desse trabalho.

Benaissa et al. (2019a) diferenciam-se dos outros trabalhos por implementar uma DT diretamente no dispositivo para classificar, em tempo real, as amostras relacionadas às classes “comer” e “ruminar”. Além disso, foi feita uma comparação com duas outras soluções: um modelo usando SVM e outro usando a solução comercial *RumiWatch*. Os autores também analisaram a relação entre a taxa de amostragem do acelerômetro e a precisão do classificador. Os resultados obtidos pela solução proposta pelos pesquisadores foram semelhantes aos do *RumiWatch*. A precisão geral foi em média 93% com o classificador SVM, 90% com DT e 91% com o dispositivo *Rumiwatch*. Phi et al. (2019) também realizaram a classificação diretamente no aparelho semelhante ao trabalho de Benaissa et al. (2019a) por meio de uma DT, porém a proposta classificou quatro classes de comportamento (em pé, deitado, comendo e andando). Esse classificador, sobretudo, foi escolhido por permitir uma fácil interpretação do processo de classificação, ajustar-se bem aos dados não lineares e exigir menos poder de processamento, o que possibilita a execução diretamente no dispositivo.

Becciolini et al. (2018) combinaram sensores (acelerômetro e GPS) embarcados em um colar para identificar comportamentos ativos (“comer”, “andar”) e inativos (“ócio”, “ruminar”) por meio do classificador LDA. Os resultados foram agrupados por raça e estação

do tempo (primavera/verão e outono). Os autores sugerem que os efeitos climáticos, assim como os hábitos específicos de cada raça, são variáveis que produzem mudanças significativas no comportamento e devem ser levadas em consideração ao se projetar um sistema de classificação do comportamento de bovinos. Por exemplo, vacas Chianina apresentaram maior nível de agitação em comparação com vacas Limousin e, durante a estação fria, houve diminuição dos níveis de atividade dos animais em relação à primavera/verão.

Riaboff et al. (2020) abordaram a sequência temporal da classificação dos comportamentos. Após a classificação, o algoritmo de Viterbi (FORNEY, 1973) foi utilizado para reavaliar as classificações temporalmente dos comportamentos obtidos pelos classificadores. RF, SVM, ADA e o *Extreme Gradient Boosting* (XGBoost) foram os classificadores usados nos experimentos. As motivações dos autores foram que a classificação dos algoritmos de aprendizagem de máquina é imperfeita e a probabilidade do estado atual do comportamento do animal depende do estado imediatamente anterior. Apesar da boa precisão obtida, a aplicação do algoritmo de Viterbi não resultou em melhorias significativas.

Dutta et al. (2015) mostraram a melhoria obtida com a utilização de combinação de classificadores. Os autores usaram os classificadores *Binary Tree* (BT), LDA, NB, *k-Nearest Neighbors* (KNN) e o *Adaptive Neuro-Fuzzy Inference System*. Nessa perspectiva, *Bagging*, *Random Subspace* e o *AdaBoost* foram os três métodos de combinação de classificadores usados.

4 MATERIAIS E MÉTODO

Para a realização dos experimentos e análises neste trabalho, a linguagem Python¹ foi escolhida devido a sua simplicidade, eficiência e disponibilidade de bibliotecas para construção de algoritmos de aprendizagem de máquina e aprendizagem profunda. Para o desenvolvimento dos scripts, a plataforma de hospedagem de *notebooks* Google Colab² foi utilizada por oferecer recursos gratuitos e o compartilhamento em conjunto com outros colaboradores. Em relação às bibliotecas, destacamos as seguintes: pandas³, para melhor manipulação nas estruturas de dados; numpy⁴, para o cálculo de matrizes multidimensionais; matplotlib⁵, para visualização dos dados; sklearn⁶, para a implementação dos classificadores de aprendizagem de máquina e cálculo das medidas de performance; keras⁷, para a implementação das CNNs e suas métricas. Por fim, todos os *notebooks* contendo as implementações dos experimentos nesta dissertação podem ser acessadas no repositório remoto no github⁸.

4.1 BASE DE DADOS

A base de dados utilizada foi construída pelo trabalho de Lomba (2015), resultado de uma parceria da Embrapa Gado de Corte e a Faculdade de Computação da Universidade Federal de Mato Grosso do Sul (FACOM-UFMS). A mesma é constituída por dados de sensores inerciais (acelerômetro, magnetômetro, giroscópio e GPS) embarcados em uma coleira. A Figura 4 mostra uma das coleiras utilizadas nos experimentos.

¹ <<https://www.python.org>>

² <<https://colab.research.google.com>>

³ <<https://pandas.pydata.org>>

⁴ <<https://numpy.org>>

⁵ <<https://matplotlib.org>>

⁶ <<https://scikit-learn.org>>

⁷ <<https://keras.io>>

⁸ Moncattle - <<https://github.com/andssuu/moncattle>>



Figura 4 – Coleira equipada com a placa junto com os sensores

Fonte: (LOMBA, 2015)

Essa solução é uma evolução da proposta de Jesus (2014) composta pela placa Arduino Fio (Figura 6(a)) a fim de integrar e gerenciar os circuitos de leitura e escrita dos dados, pelo *data logger* OpenLog (Figura 6(b)) da SparkFun para o armazenamento dos dados capturados pelos sensores em um micro cartão SD, por um módulo Vênus GPS (Figura 6(c)), também da SparkFun, para obter os dados de posicionamento e pelo dispositivo MiniIMU 9DOF v3 (Figura 6(d)) encarregado pela captura dos dados relacionados à movimentação providos pelos seguintes sensores: acelerômetro, giroscópio e magnetômetro, onde cada um deles possui 3 eixos (x, y e z) e foram configurados com a taxa de frequência de 1Hz, ou seja, a cada um segundo, uma captura era realizada.

Houve uma preocupação, da partes dos pesquisadores, acerca da posição da coleira no corpo do animal a fim de padronizar o referencial dos eixos dos sensores e não causar inconstâncias durante o treinamento dos algoritmos, uma vez que a localização da coleira influencia diretamente os valores dos eixos dos sensores de movimento. A Figura 5 ilustra a orientação dos eixos do acelerômetro de acordo com o seguinte padrão de vestimenta: o lado em que o cinto é ajustado com a fivela precisa ficar ao lado esquerdo do animal e a caixa contendo os sensores deve ficar abaixo do pescoço.

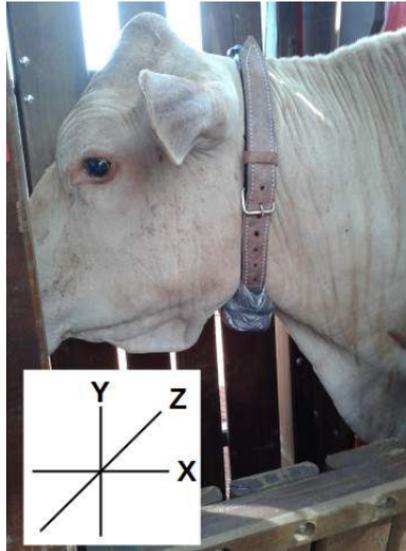
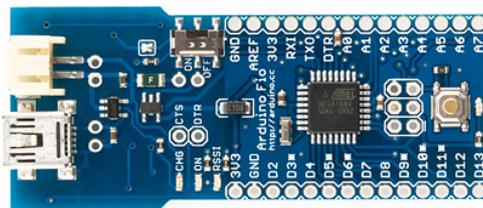
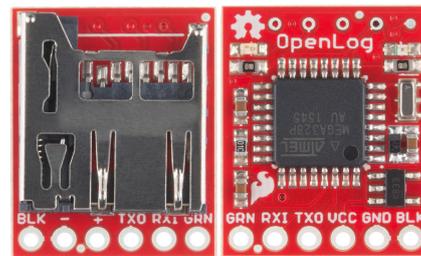


Figura 5 – Animal vestido com a coleira de acordo com o padrão
Fonte: (LOMBA, 2015)



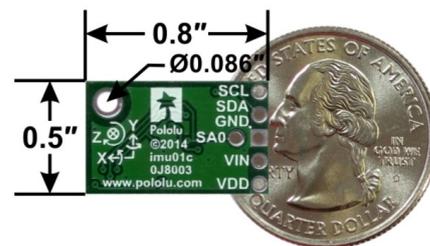
(a) Fonte: Arduino¹



(b) Fonte: Sparkfun²



(c) Fonte: Sparkfun³



(d) Fonte: Pololu⁴

Figura 6 – Componentes utilizados na solução de monitoramento (a) Arduino Fio (b) OpenLog (c) Vênus GPS (d) MinIMU-9 v3

¹ Arduino Fio - <<https://www.arduino.cc/en/pmwiki.php?n=Main/ArduinoBoardFio>>

² SparkFun OpenLog - <<https://www.sparkfun.com/products/13712>>

³ SparkFun Venus GPS with SMA Connector - <<https://www.sparkfun.com/products/retired/11058>>

⁴ MinIMU-9 v3 Gyro, Accelerometer, and Compass (L3GD20H and LSM303D Carrier) - <<https://www.pololu.com/product/2468>>

Os experimentos foram realizados em dois períodos (14 a 28 de maio e 25 de junho a 06 de julho de 2015) na área de pasto do Mangueiro Digital, laboratório dedicado a pesquisas de Pecuária de Precisão da Embrapa Gado de Corte. Apesar da disponibilidade de oito animais mantidos aos cuidados da Embrapa para os experimentos, em cada período de experimento, apenas dois animais foram selecionados aleatoriamente e equipados com os colares.

A Tabela 1 contém as características dos animais que foram usados durante os experimentos. Cada um deles era identificado com um código que estava marcado na sua pele. Entretanto, percebe-se, na Tabela 1, que há o registro de apenas três animais. Isso ocorreu devido ao animal 6998 ter participado dos dois experimentos, porém de maneira aleatória, uma vez que a escolha dos animais era feita com base nos dois primeiros animais que entravam no tronco. Outro ponto a se destacar foi a cautela dos autores de considerar um período de 24 horas para a adaptação dos animais com as coleiras antes de realizar a efetiva captura dos dados. Essa prática também foi adotada por Benaissa et al. (2019a).

Tabela 1 – Informações dos animais utilizados nos experimentos

ID	RAÇA	PELAGEM	CATEGORIA	NASCIMENTO
6998	CARACU	VERMELHA	NOVILHA	19/09/2013
7014	CARACU	VERMELHA	NOVILHA	23/09/2013
7135	CARACU	BAIA	NOVILHA	01/11/2013

Tal decisão, por utilizar apenas dois animais, ocorreu após a explicação dos profissionais da Embrapa a respeito da complexidade associada às atividades do processo de rotulação dos comportamentos dos animais, posto que as anotações eram realizadas por meio da observação visual *in loco*. Diante disso, como uma medida paliativa a fim de reduzir o erro e aumentar a precisão da rotulação, uma versão adaptada do aplicativo móvel desenvolvido por Jesus (2014) foi utilizada no registro das classes de comportamentos. A Figura 7 ilustra os comportamentos levados em consideração durante o processo de rotulação dos dados. Houve uma divisão dos comportamentos em primários e secundários. Os primeiros eram compostos pelos comportamentos “Pastando/Procurando”, “Andando”, “Em pé” e “Deitado”. A seguir, a definição cada um deles:

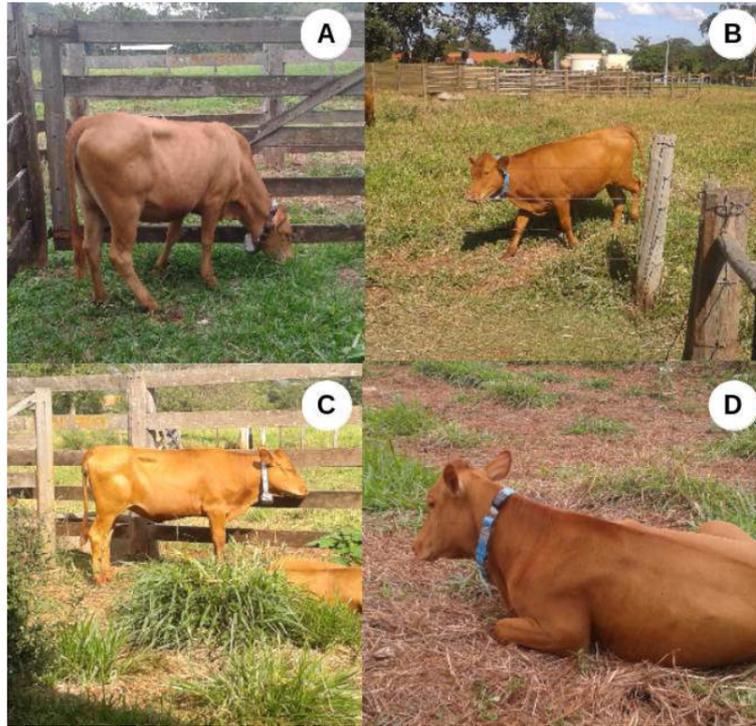


Figura 7 – Classes dos comportamentos de bovinos levadas em consideração na rotulação:
(A) Pastando, (B) Andando, (C) Em pé e (D) Deitado

Fonte: (LOMBA, 2015)

- Pastando/Procurando: o animal está sobre as quatro patas com a cabeça baixa procurando ou mastigando o capim. O animal pode ou não estar em movimento, uma vez que ele pode estar se deslocando à procura de capim. Nesta dissertação, por simplicidade, consideramos essa classe apenas como “Pastando”;
- Andando: o animal também está sobre as quatro patas, porém com o pescoço reto (apontando o focinho para frente) e se deslocando pela área de pasto;
- Em Pé: o animal está sobre as quatro patas com a cabeça erguida e não há deslocamento;
- Deitado: o animal não está apoiado em nenhum dos membros, podendo ser em decúbito lateral ou ventral.

Já os secundários consistem em uma ramificação dos comportamentos “Em Pé” e “Deitado” adicionando o comportamento ruminar, o qual ocorre quando o animal está com a mandíbula em movimento. Logo, duas novas classes foram adicionadas: “Ruminando Em Pé” e “Ruminando Deitado”. Ao contrário de Lomba (2015), esta dissertação leva em consideração a classificação simultânea de todas as seis classes de comportamentos.

Por fim, a Tabela 2 mostra algumas amostras da estrutura final da base de dados

após a captura e sincronização da rotulação dos dados. No total, a base é composta por 13088 amostras. Percebe-se, na Tabela 2, que dos dados obtidos pelo GPS, apenas a velocidade de deslocamento (“velDesloc”) foi utilizada nos experimentos.

Tabela 2 – Estrutura da base de dados

accX	accY	accZ	magX	magY	magZ	girX	girY	girZ	velDesloc	classe
1157	9002	-12840	-460	1575	-205	243	195	-466	0	EmPe-Ruminando
1226	9986	-12546	-463	1585	-264	-492	2209	-2189	0	EmPe-Ruminando
-5804	12889	-7345	360	318	-1563	207	263	34	0	Deitado-Parado
-1957	7320	-12335	972	1353	-171	-3845	5742	-5178	2.6	Andando
280	8976	-10232	653	-207	-1361	-1948	3395	-2120	2.7	EmPe-Ruminando

accX: eixo x do acelerômetro.
accY: eixo y do acelerômetro.
accZ: eixo z do acelerômetro.
magX: eixo x do acelerômetro.
magY: eixo y do acelerômetro.
magZ: eixo z do acelerômetro.
girX: eixo x do acelerômetro.
girY: eixo y do acelerômetro.
girZ: eixo z do acelerômetro.
velDesloc: velocidade de deslocamento.

4.2 MÉTODO

A Figura 8 mostra uma síntese da heurística utilizada neste trabalho para a classificação do comportamento de bovinos. Inicialmente, realizou-se uma divisão da base mediante a técnica *holdout*, a qual consiste em dividir o conjunto de dados em duas partes mutuamente exclusivas, com 60% (7852 amostras) para a base de treino e 40% (5236 amostras) para a base de teste. A escolha dessas porcentagens para composição das bases se deu após uma análise empírica dos resultados com o objetivo de evitar uma variância significativa em relação às métricas da base de teste.

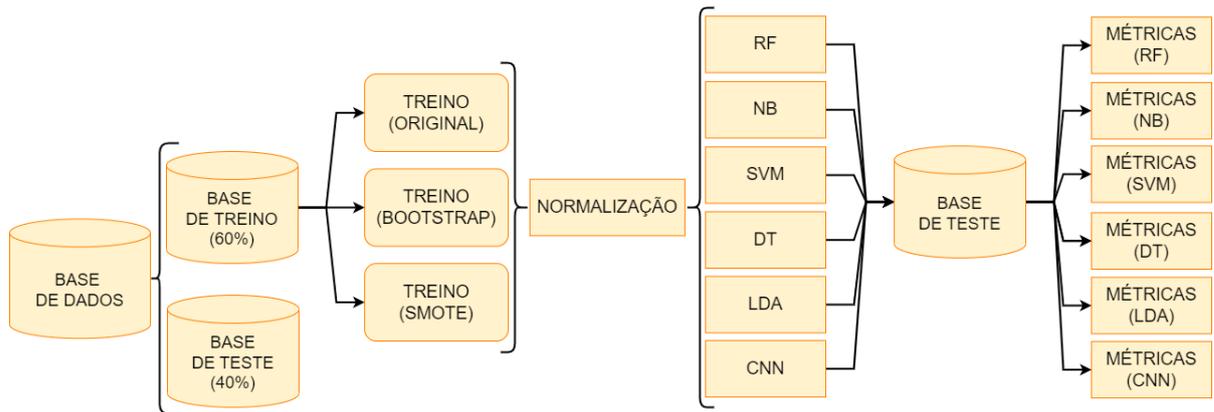


Figura 8 – Fluxograma do método

Uma vez que a base de dados estava desbalanceada e a divisão dos dados foi realizada de maneira estratificada, conseqüentemente, a base de treinamento e teste também estavam desbalanceadas. Conforme mostra a Figura 9(a), nota-se os níveis de desbalanceamento das classes para a base de treinamento. Diante disso, houve a aplicação das técnicas de reamostragem *bootstrap* e SMOTE resultando em três cenários para os dados de treinamento: base original, base *bootstrap* e base SMOTE. Nesta dissertação, consideramos o termo “original” para a base com os dados sem reamostragem. A Figura 9(b) mostra a distribuição das classes depois do processo de reamostragem. Em seguida, aplicou-se a normalização nos dados de treinamento para cada um dos cenários pela Equação 2.1.

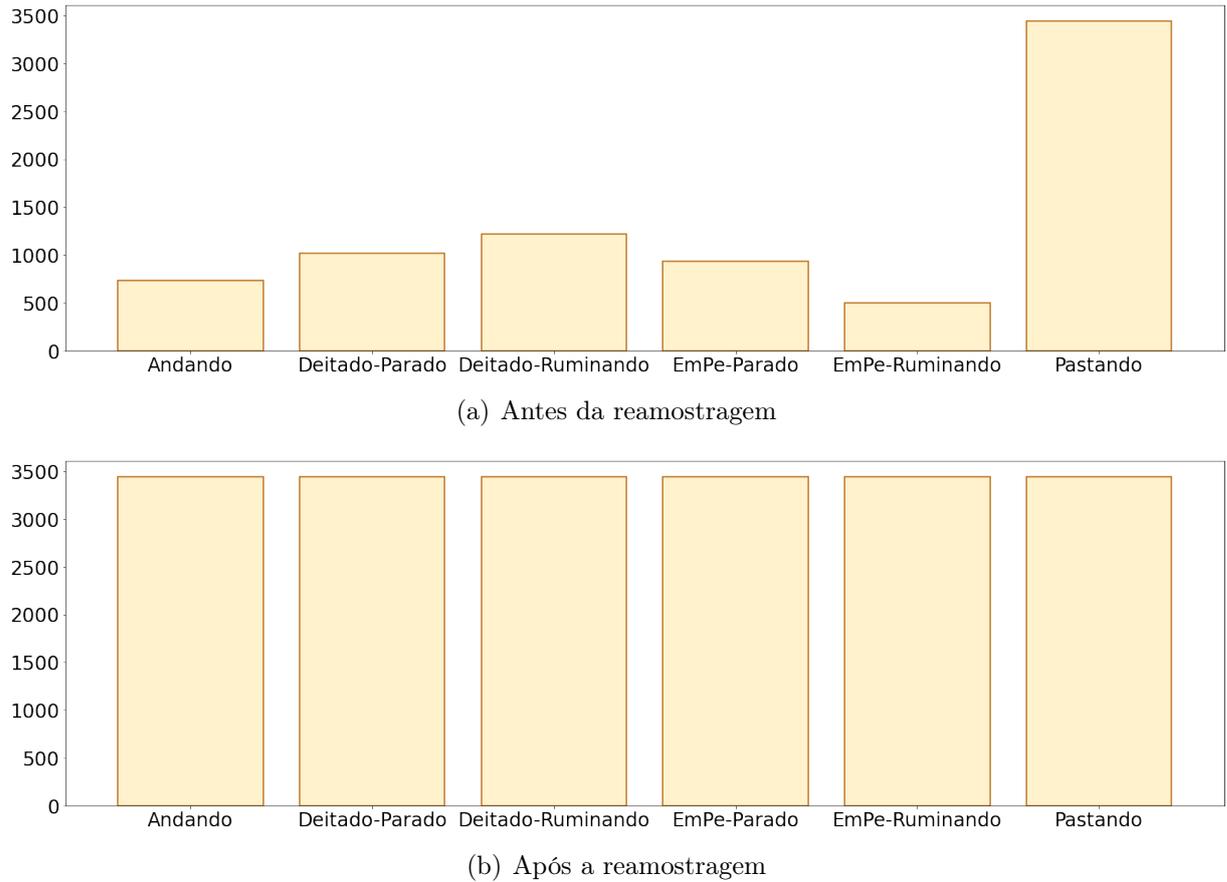


Figura 9 – Distribuição dos dados em relação às classes de comportamentos (a) antes da reamostragem e (b) após a reamostragem

Por fim, após essas etapas de tratamento nos dados, houve o efetivo treinamento dos algoritmos de classificação. Além das CNNs propostas nesta dissertação, as quais serão detalhadas no Capítulo 5, outras cinco CNNs (Lenet, AlexNet, MiniVGG, Resnet e Mobilenet) e mais cinco classificadores de aprendizagem de máquina (RF, NB, SVM, DT, LDA) foram utilizados como comparativo. Cada um desses classificadores foi avaliado de acordo com as seguintes métricas em relação à base de teste: acurácia, precisão, sensibilidade e f1-score. Todas essas medidas foram abordadas na Subseção 2.3.4.

5 ARQUITETURAS

Este capítulo descreve as duas arquiteturas otimizadas das CNNs propostas para a classificação do comportamento de bovinos. Como a temporalidade linear não é uma propriedade obedecida nos dados presentes na base de dados, devido ao processo de rotulação acontecer em diferentes momentos e de forma não constante, além de levar em consideração diferentes coleiras em diferentes animais e em diferentes momentos, buscou-se analisar o problema da classificação do comportamento de bovinos, além da técnica CONV1D, pela perspectiva espacial da relação dos dados dos sensores por meio da CONV2D. Conforme mostra a Figura 10, a estrutura dos padrões de entrada é composta apenas por um vetor com dez dimensões. Desta forma, além de possuir baixa dimensão, essa estrutura acarreta em um problema durante a extração de características por meio da aplicação dos filtros de convolução: induzir os filtros a levarem em consideração apenas a relação de alguns dados dos sensores entre si.



Figura 10 – Vetor de características originário

A Figura 11 ilustra esse problema. Se pensarmos na execução de um filtro de convolução de tamanho 3x3, por exemplo, perceberemos que não haveria a extração de características que contemple a relação entre os dados do acelerômetro com os do magnetômetro ou com a velocidade de deslocamento. Portanto, essa estrutura não é interessante para o modelo, pois deixaria de levar em consideração relações entre dados que poderiam ser significativas para a discriminação de um determinado comportamento. A Figura 11 também mostra o uso da técnica *zero padding*, a qual consiste em preencher com zeros a borda do vetor a fim de possibilitar a execução dos filtros. Pelo seu amplo uso e simplicidade, essa técnica foi utilizada nas arquiteturas durante os experimentos para o preenchimento da borda dos vetores de características.

Diante disso, foram realizadas transformações para a expansão dos padrões de entrada por meio de duas abordagens a fim de possibilitar uma extração de características mais eficiente e, por consequência, uma melhor performance do modelo: a expansão do vetor de características por combinação e a expansão do vetor de características pelo uso de

uma matriz de ativações. Isso decorre devido ao vetor de características expandido permitir com que os modelos considerem várias dimensões que não são viáveis quando aplicados em dados de baixa dimensão (JUNG et al., 2021). Cada uma dessas duas abordagens são detalhadas nas seções seguintes.

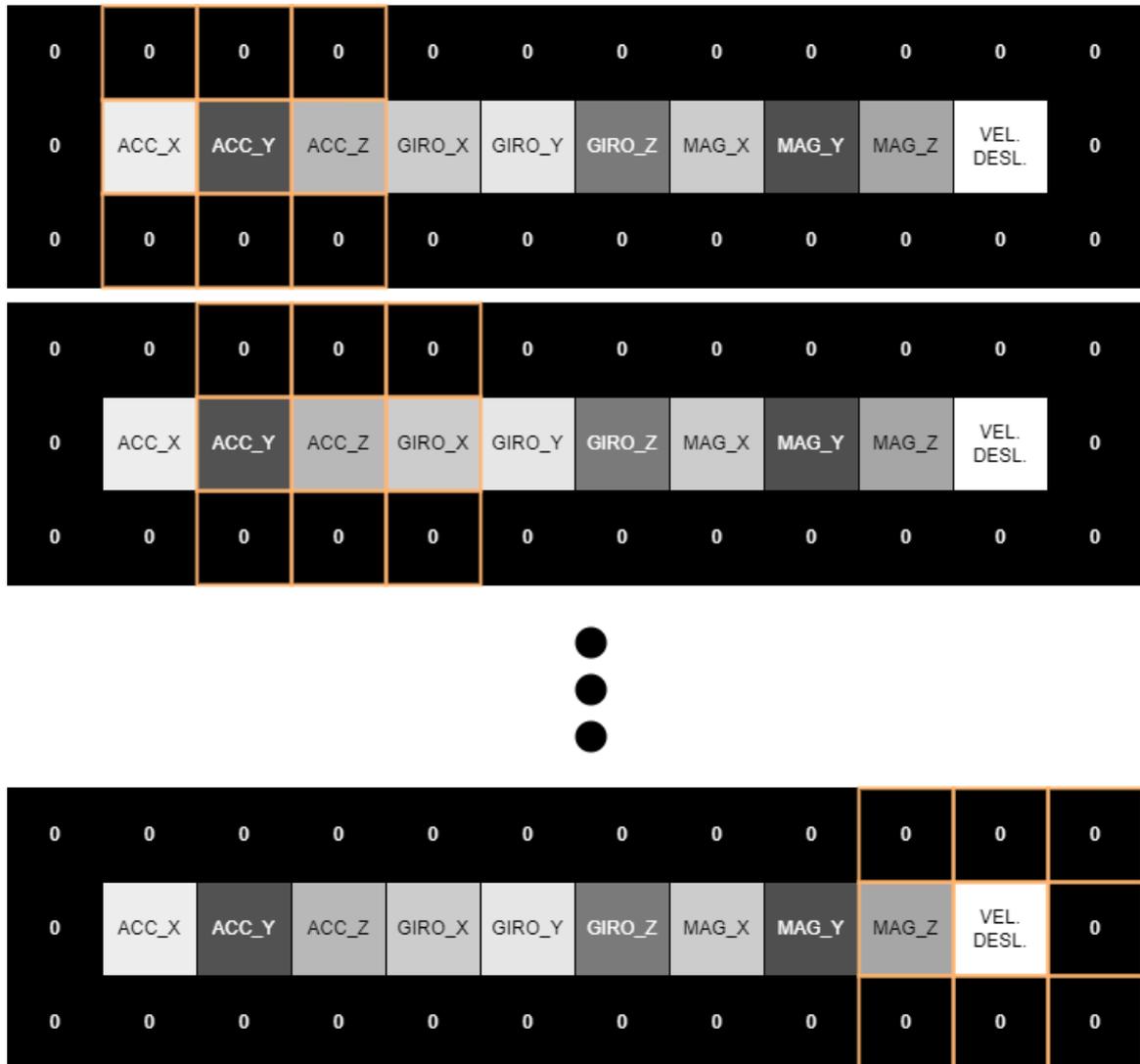


Figura 11 – Filtro de convolução com dimensão 3x3 sobre o vetor de características original

5.1 EXPANSÃO DAS CARACTERÍSTICAS POR COMBINAÇÃO

Nesse cenário, adicionou-se ao vetor de característica original os mesmos dados, mas em posições diferentes, fundamentado na seguinte regra: cada nova linha adicionada ao vetor de característica é formada pelo deslocamento de uma posição à direita dos dados, onde o último elemento, após o deslocamento, ficará na primeira posição. Essas

operações acontecem até o primeiro elemento do vetor de característica original (eixo x do acelerômetro) ocupar a última posição da linha adicionada. Essa estratégia é uma forma de fazer com que a aplicação dos filtros da CNN consiga extrair características que levem em consideração um maior número de relações entre os dados dos sensores. Conforme mostra a Figura 12, o resultado dessa expansão é um vetor de característica representado por uma matriz de dimensão 10x10.

ACC_X	ACC_Y	ACC_Z	GIRO_X	GIRO_Y	GIRO_Z	MAG_X	MAG_Y	MAG_Z	VEL. DESL.
VEL. DESL.	ACC_X	ACC_Y	ACC_Z	GIRO_X	GIRO_Y	GIRO_Z	MAG_X	MAG_Y	MAG_Z
MAG_Z	VEL. DESL.	ACC_X	ACC_Y	ACC_Z	GIRO_X	GIRO_Y	GIRO_Z	MAG_X	MAG_Y
MAG_Y	MAG_Z	VEL. DESL.	ACC_X	ACC_Y	ACC_Z	GIRO_X	GIRO_Y	GIRO_Z	MAG_X
MAG_X	MAG_Y	MAG_Z	VEL. DESL.	ACC_X	ACC_Y	ACC_Z	GIRO_X	GIRO_Y	GIRO_Z
GIRO_Z	MAG_X	MAG_Y	MAG_Z	VEL. DESL.	ACC_X	ACC_Y	ACC_Z	GIRO_X	GIRO_Y
GIRO_Y	GIRO_Z	MAG_X	MAG_Y	MAG_Z	VEL. DESL.	ACC_X	ACC_Y	ACC_Z	GIRO_X
GIRO_X	GIRO_Y	GIRO_Z	MAG_X	MAG_Y	MAG_Z	VEL. DESL.	ACC_X	ACC_Y	ACC_Z
ACC_Z	GIRO_X	GIRO_Y	GIRO_Z	MAG_X	MAG_Y	MAG_Z	VEL. DESL.	ACC_X	ACC_Y
ACC_Y	ACC_Z	GIRO_X	GIRO_Y	GIRO_Z	MAG_X	MAG_Y	MAG_Z	VEL. DESL.	ACC_X

Figura 12 – Formação do vetor de características por combinação

Após essa transformação no vetor originário, as amostras são processadas pela arquitetura ilustrada na Figura 13. Essa arquitetura foi inspirada na Lenet-5 proposta por LeCun et al. (1998). Entretanto, nota-se a adição de camadas de normalização após a execução dos filtros como uma forma de estabilizar e reduzir o número de épocas necessárias para a convergência da rede. Além disso, houve a necessidade da aplicação das técnicas *dropout* e regularização L1 a fim de reduzir o *overfitting* dos modelos, as quais foram apresentadas na Subseção 2.3.5 e serão discutidas em mais detalhes na Seção 5.3. Após a extração de características pelos filtros, há duas camadas totalmente conectadas de neurônios com a função de ativação ReLU. Por fim, a última camada da rede é a responsável pela classificação propriamente dita, consistindo em seis neurônios com a

função *softmax*, uma vez que são seis classes de comportamentos a serem classificadas, em que cada uma delas é responsável por informar uma probabilidade da amostra pertencer a uma determinada classe.

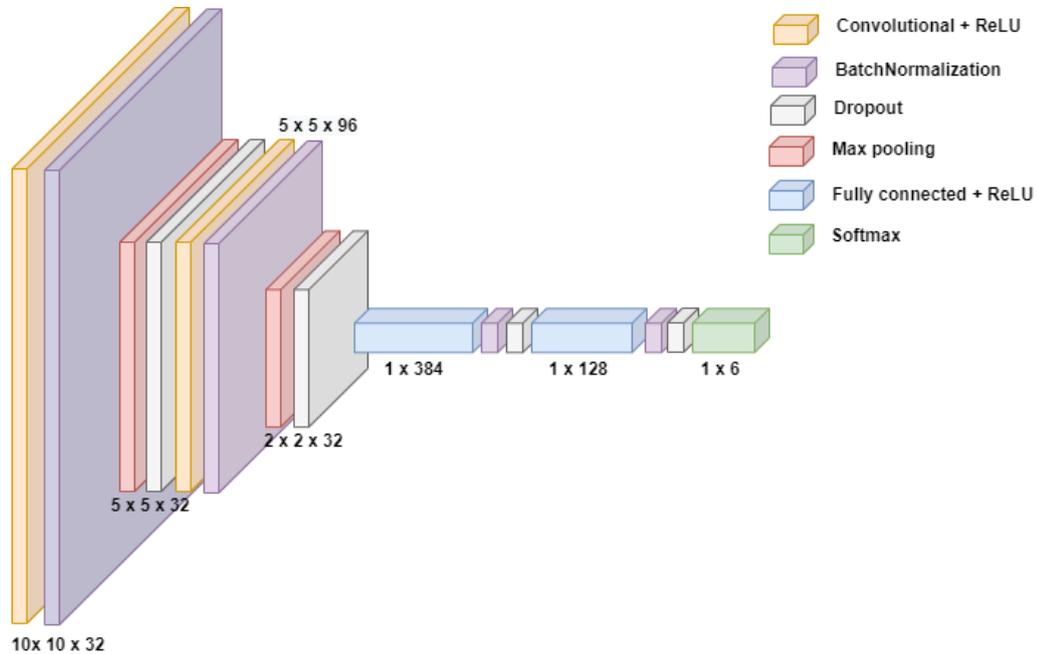


Figura 13 – Arquitetura CNN por combinação

5.2 EXPANSÃO DAS CARACTERÍSTICAS PELA MATRIZ DE ATIVAÇÃO

Em virtude do vetor de característica possuir baixa dimensão, buscou-se uma estratégia que expandisse as dimensões para o mesmo tamanho das entradas da arquitetura proposta por LeCun et al. (1998). Assim, expandiu-se o vetor de características por meio de uma matriz de ativações integrada na própria arquitetura da rede. A Figura 14 ilustra a arquitetura dessa abordagem. Inicialmente, aplica-se uma camada densa composta por neurônios com a função de ativação ReLU. Após essas ativações, há uma normalização dos dados e uma remodelagem no vetor de características composto por 1024 dimensões para uma matriz com dimensão 32x32. Assim, a extração de características se dará pelas convoluções dos filtros da CNN em torno das ativações dos neurônios presentes nessa matriz baseadas em cada padrão de entrada.

A essência dessa abordagem se concentra na seguinte ideia: após a convergência dos pesos das camadas densas junto com os valores dos filtros de convolução, os padrões

de uma dada classe ativarão determinados neurônios espalhados em regiões específicas na matriz. Ressalta-se que um padrão de entrada pode ativar neurônios em mais de uma região nessa matriz, sendo assim, não há uma relação obrigatória entre as ativações com a espacialidade de vizinhança entre os neurônios na matriz. A Figura 15 ilustra algumas representações possíveis das matrizes após as ativações para duas amostras de diferentes classes. Todo o restante dessa arquitetura é semelhante ao já apresentado anteriormente na Seção 5.1.

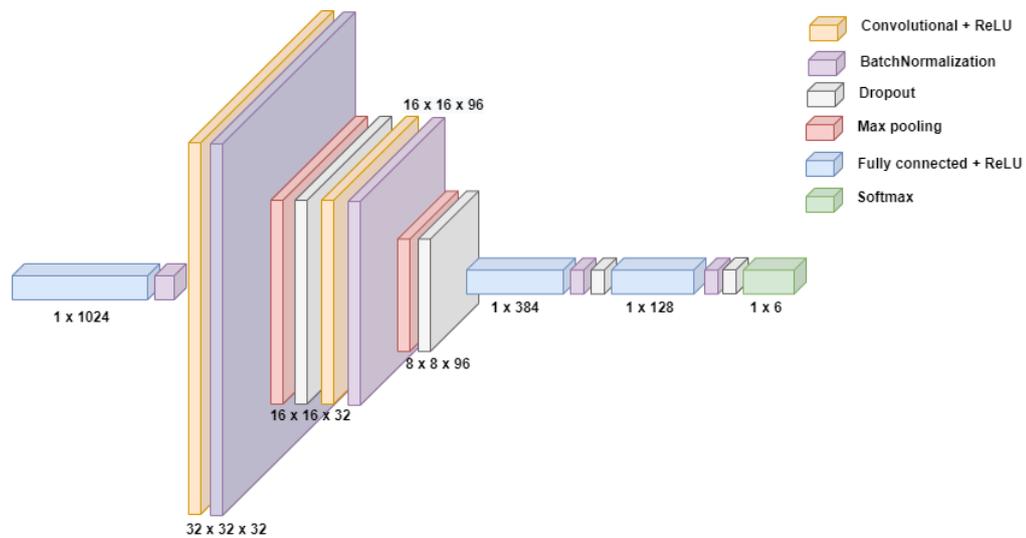


Figura 14 – Arquitetura CNN por matriz de ativação

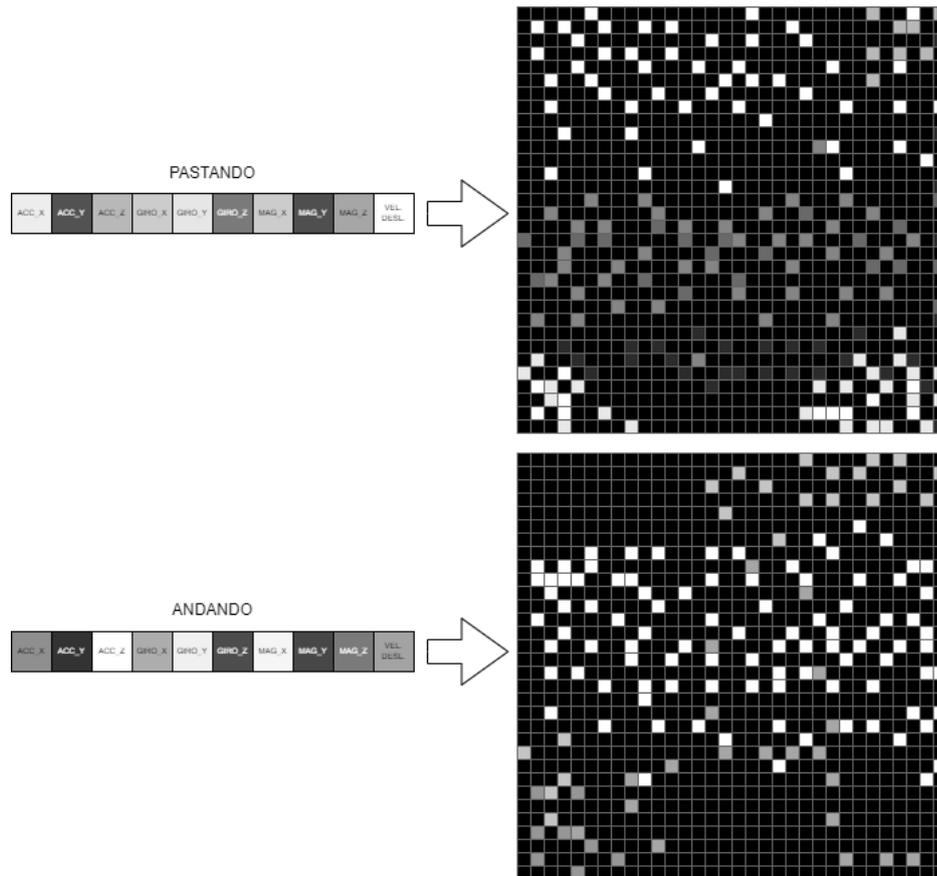


Figura 15 – Ativações na matriz de neurônios para padrões de classes diferentes

5.3 OTIMIZAÇÃO

O projeto de uma CNN requer uma série de tomadas de decisões, assim, para definir a arquitetura e escolha dos hiperparâmetros da rede, foi realizada uma outra divisão novamente pela técnica *holdout*, além da divisão inicial das bases de treino e teste, sobre a base de treino original consistindo uma parte dos dados para o efetivo treinamento dos modelos composta por 80% das amostras e outra parte para a validação composta por 20% dos dados. Em especial, essa divisão é uma maneira fundamental para identificação e redução de problemas habituais nos modelos, tais como o *overfitting*, pois as medidas de avaliação são tomadas em relação aos dados não apresentados durante o treinamento a fim de reduzir o viés e projetar um cenário mais real e confiável para o desenvolvimento dos modelos. A Figura 16 ilustra essa divisão junto com o fluxo seguido para treinamento e análise da performance dos modelos em relação ao conjunto de validação.

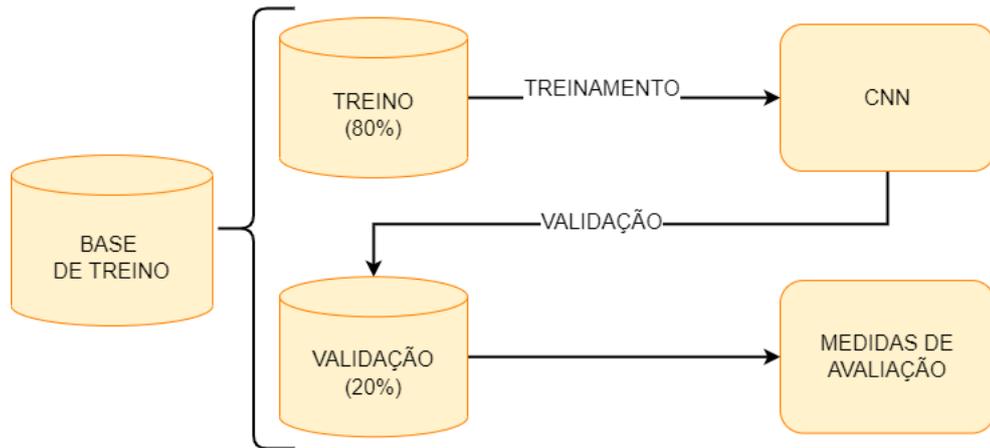


Figura 16 – Fluxograma para a escolha dos hiperparâmetros da CNN

Inicialmente, as arquiteturas das redes foram inspiradas na rede proposta por LeCun et al. (1998). Entretanto, ao invés de utilizar camada de subamostragem pela função da média, escolheu-se a função máximo a fim de obter uma convergência mais rápida da rede. Também houve uma adaptação em relação ao tamanho dos filtros devido ao vetor de características apresentar uma menor dimensão. A Figura 17 ilustra essas arquiteturas iniciais.

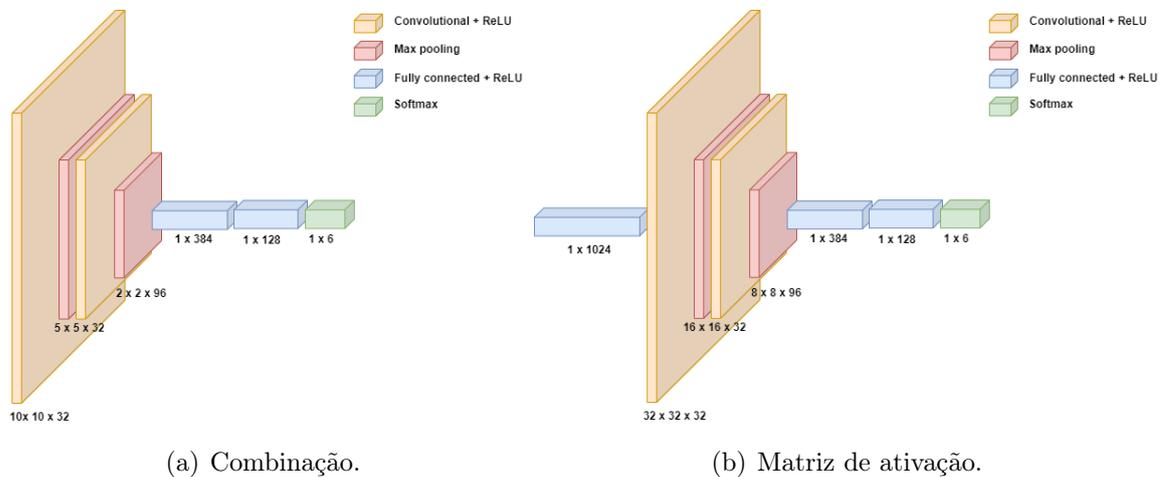


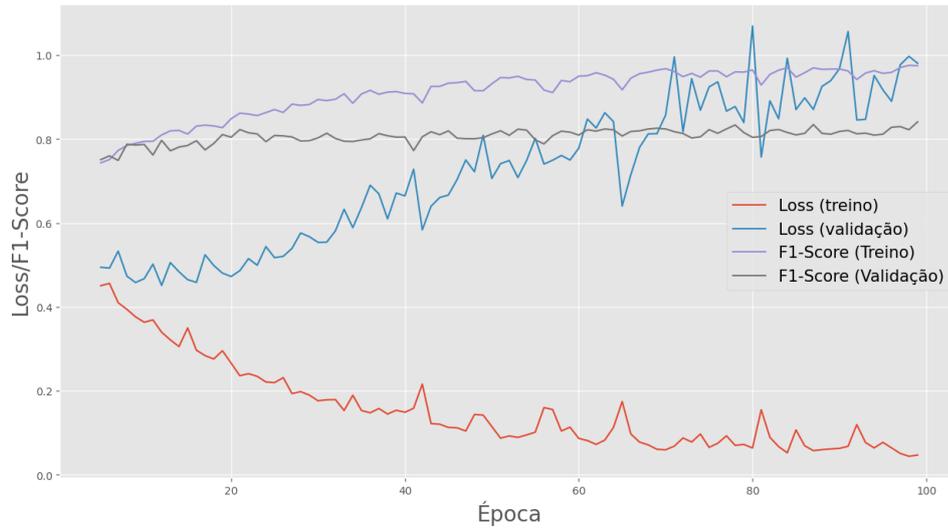
Figura 17 – Arquiteturas iniciais das CNNs (a) combinação e (b) matriz de ativação

Devido ao desbalanceamento dos dados, a medida f1-score foi escolhida para a avaliação da performance dos modelos por considerar a média harmônica entre a precisão e a sensibilidade e, conseqüentemente, representar uma informação mais confiável para a tomada de decisão nesse contexto, ao contrário de outras medidas, como por exemplo a acurácia.

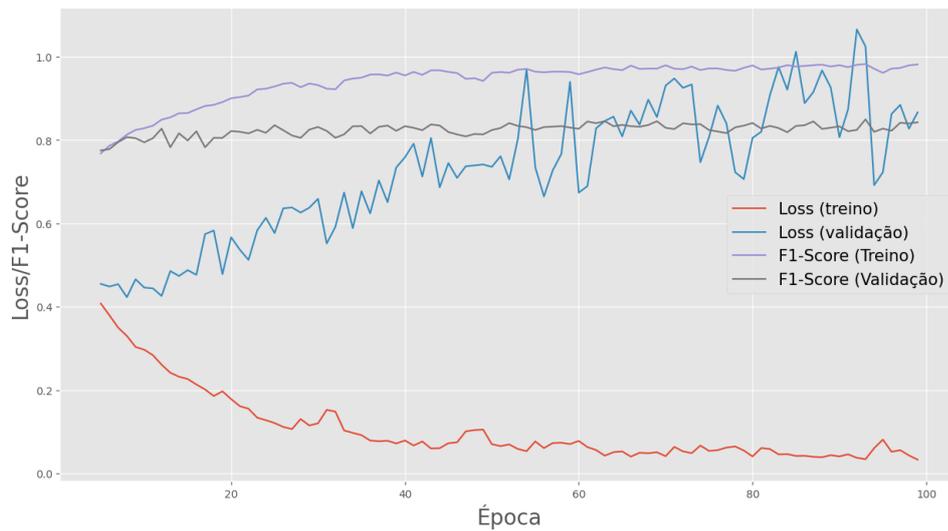
A Figura 18 mostra os gráficos de desempenho com as medidas *loss* e f1-score das CNNs em relação ao cenário da base original sobre os dados de validação ao decorrer das épocas (iterações durante o treinamento). Nota-se a presença de *overfitting* com início em torno da décima época em ambas abordagens.

É fundamental ressaltar o cuidado ao interpretar os valores obtidos nos cenários com reamostragem (bases *bootstrap* e SMOTE), pois podem acabar resultando em expectativas não ideais sobre a base de teste. Isso se deve pela própria formação dos dados de validação nesses cenários. Com *bootstrap*, a base de validação tende a possuir um grau maior de similaridade com a base de treino devido a presença de algumas amostras duplicadas tanto na base de treino quanto nos dados de validação e isso faz com que o modelo se ajuste bem em ambas as bases. Nos dados com reamostragem SMOTE, essa situação também se repete, mas em menor proporção, pois a geração de novas amostras sintéticas pela técnica SMOTE promove uma maior variabilidade entre dados da mesma classe. Diante disso, apesar da construção dos gráficos e das medidas de avaliação para os cenários com reamostragem, toda a análise para verificação da remoção do *overfitting* foi focada no cenário com a base original.

Ainda com relação aos gráficos da Figura 18, outro ponto a se destacar é a convergência mais rápida da abordagem pela matriz de ativação em comparação com a abordagem por combinação. Isso fica perceptível se analisarmos as inclinações das linhas que representam as medidas f1-score em relação à base de validação. Se tomarmos a vigésima época como referência, perceberíamos que a CNN com matriz de ativação obteve melhor performance do que a CNN por combinação. Entretanto, essas diferenças tendem a diminuir à medida que aumentamos o número de épocas.



(a) CNN inicial com expansão por combinação

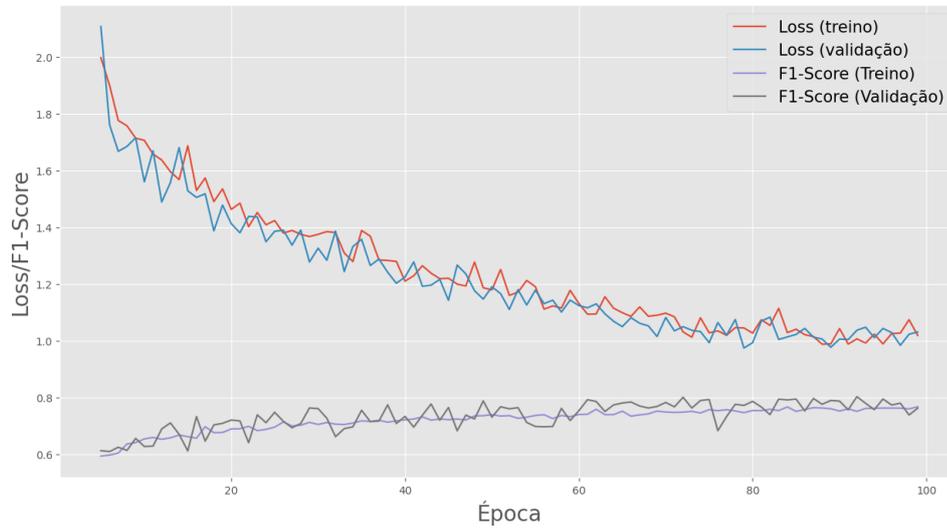


(b) CNN inicial com expansão pela matriz de ativação

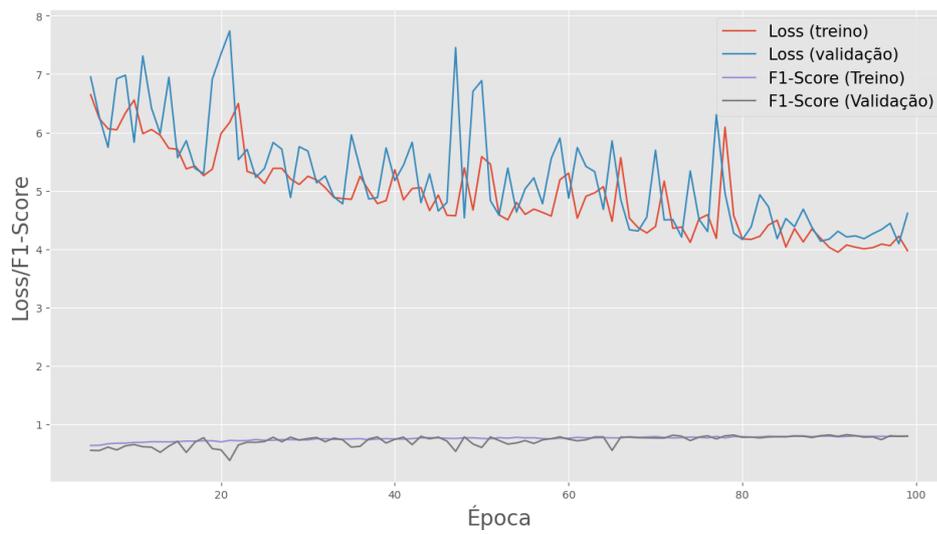
Figura 18 – Performances das CNNs em relação à base de validação no cenário com a base original

Dessa forma, diante da presença de *overfitting* nos modelos, foram aplicadas a técnica de regularização L1 (com o fator de regularização igual à 0.01) e o *dropout* (com taxa de descarte de 20%) nas arquiteturas a fim de mitigar tal situação. A Figura 19 mostra os desempenhos das CNNs após a otimização em relação ao cenário com a base original sobre os dados de validação. Ficam nítidas as diferenças com os modelos anteriores provendo uma sincronia entre as medidas f1-score e o *loss* em relação aos dados de treino e validação. Além disso, percebe-se que, mesmo após os ajustes das redes, a abordagem com a matriz de ativações ainda continuou convergindo mais rápido do que a abordagem por combinação.

Por fim, no Apêndice A, as Tabelas A.1 e A.2 contêm mais detalhes sobre as camadas das arquiteturas otimizadas das CNNs em ambas abordagens.



(a) CNN otimizada com expansão por combinação



(b) CNN otimizada com expansão pela matriz de ativação

Figura 19 – Performances das CNNs otimizadas em relação à base de validação no cenário com a base original

6 RESULTADOS E DISCUSSÕES

Neste capítulo, são reportadas as performances, em relação à base de teste, das CNNs junto com as dos classificadores de aprendizagem de máquina. A Seção 6.1 dá ênfase nos resultados e nas análises dos desempenhos entre as CNNs baseadas nas duas abordagens de expansão do vetor de características e a CNN com os dados não expandidos, por meio da técnica de convolução CONV1D, sobre cada uma das seis classes de comportamentos. Na Seção 6.2, são apresentadas as médias das métricas de todos os modelos utilizados nos experimentos e realizadas algumas comparações entre os principais resultados das CNNs com os algoritmos de aprendizagem de máquina. Na Seção 6.3, são identificados e discutidos alguns tópicos relevantes oriundos das decisões de projeto e dos resultados obtidos.

6.1 COMBINAÇÃO X MATRIZ DE ATIVAÇÃO X CONV1D

Esta seção enfatiza os resultados da CNN com CONV1D e das CNNs por combinação e pela matriz de ativações em relação a base de teste nos três cenários dos dados de treinamento, pois, busca-se analisar os resultados dos modelos com essas diferentes abordagens, além do mais, diferentemente dos outros algoritmos, essas redes ainda não foram muito empregadas na classificação do comportamento de bovinos. Para a análise das CNNs, decidiu-se utilizar as medidas de avaliação de precisão, sensibilidade e f1-score, as quais fornecem informações suficientes para realizar comparações entre os modelos. As Tabelas 3, 4 e 5 representam, respectivamente, as medidas de desempenho das abordagens por combinação (colunas COMB), por matriz de ativação (colunas MA) e por CONV1D (colunas 1D) para cada um dos cenários da base original, *bootstrap* e SMOTE.

Tabela 3 – Medidas de avaliação por classe de comportamento das CNNs em relação à base de teste - cenário com base de treino original

	Precisão (%)			Sensibilidade (%)			F1-Score (%)		
	1D	COMB	MA	1D	COMB	MA	1D	COMB	MA
Pastando	94.65	94.98	93.19	96.95	96.55	97.82	95.74	95.76	95.45
Deitado-Ruminando	76.79	78.01	78.49	81.70	78.01	75.31	79.17	78.01	76.87
Deitado-Parado	81.27	69.39	72.44	67.55	74.89	71.81	73.78	72.03	72.12
EmPe-Parado	69.79	71.16	68.70	69.23	66.83	70.35	69.51	68.93	69.52
Andando	90.49	92.44	93.64	91.04	89.61	86.97	90.76	91.00	90.18
EmPe-Ruminando	58.24	59.52	67.63	59.46	52.55	56.46	58.84	55.82	61.54
Média	78.52	77.58	79.01	77.65	76.4	76.45	77.97	76.92	77.61

Tabela 4 – Medidas de avaliação por classe de comportamento das CNNs em relação à base de teste - cenário com base de treino *bootstrap*

	Precisão (%)			Sensibilidade (%)			F1-Score (%)		
	1D	COMB	MA	1D	COMB	MA	1D	COMB	MA
Pastando	96.45	95.78	95.72	94.90	96.03	95.64	95.67	95.91	95.68
Deitado-Ruminando	84.35	80.83	83.87	74.82	81.82	85.63	79.30	81.32	84.74
Deitado-Parado	72.16	76.95	81.58	78.41	75.48	77.39	75.16	76.20	79.43
EmPe-Parado	73.64	74.38	79.29	69.39	72.12	68.11	71.45	73.23	73.28
Andando	87.38	88.02	87.67	95.93	94.30	94.09	91.46	91.05	90.77
EmPe-Ruminando	61.64	70.38	65.93	72.37	66.37	80.18	66.57	68.32	72.36
Média	79.27	81.05	82.34	80.97	81.02	83.50	79.93	81	82.71

Tabela 5 – Medidas de avaliação por classe de comportamento das CNNs em relação à base de teste - cenário com base de treino SMOTE

	Precisão (%)			Sensibilidade (%)			F1-Score (%)		
	1D	COMB	MA	1D	COMB	MA	1D	COMB	MA
Pastando	96.76	96.33	94.73	93.76	93.85	96.34	95.24	95.07	95.52
Deitado-Ruminando	77.52	78.69	86.47	86	87.10	82.43	81.54	82.68	84.40
Deitado-Parado	81.33	82.86	80.95	68.43	68.87	80.47	74.32	75.22	80.71
EmPe-Parado	70.67	68.77	76.13	70.67	75.16	73.08	70.67	71.82	74.57
Andando	86.57	89.60	91.50	94.50	91.24	85.54	90.36	90.41	88.42
EmPe-Ruminando	62.96	69.12	65.39	71.47	73.27	77.18	66.95	71.14	70.80
Média	79.30	80.9	82.53	80.81	81.58	82.5	79.85	81.06	82.4

De início, foram analisadas as métricas dos modelos baseados nas CNNs a fim de verificar qual das abordagens obteve os melhores resultados e, além disso, investigar as semelhanças e diferenças nas classificações em cada um dos três cenários das bases de treinamento. Comparando os resultados individualmente das CNNs de acordo com as médias de F1-Score para cada cenário representadas nas Tabelas 3, 4 e 5, é interessante destacar que, no geral, o uso das técnicas de reamostragem *bootstrap* e SMOTE sucedeu em melhorias nos modelos em relação ao cenário com a base original, tendo as abordagens por combinação e pela matriz de ativação obtido as melhorias mais significativas, uma vez que a diferença entre os cenários com CONV1D ficaram abaixo de 2%, enquanto que para as outras abordagens, essa diferença ficou acima de 4%.

Analisando os resultados individuais das classes de comportamentos para cada cenário, nota-se que houve melhoria em relação a quase todas as classes de comportamentos, menos alguns casos para as classes “Pastando” e “Andando”, entretanto as diferenças acabaram não sendo significativas ficando em torno de 1%. Esse comportamento apesar de ser pouco significativo, é compreensível, pois nos casos com reamostragem há uma tentativa de remoção do viés dos modelos para a classe majoritária, nesse caso, a classe “Pastando”.

Em especial, o aumento da performance nas classificações em relação à classe minoritária aconteceu em todos os cenários e torna-se um ponto convincente na justificativa do uso das técnicas de reamostragem. Por exemplo, a medida f1-score relacionada à classe “EmPe-Ruminando”, a qual possui a menor quantidade de padrões na base de teste com

333 amostras, apresentou um aumento em torno de 8% para a abordagem CONV1D, 14% para a abordagem por combinação e 10% para a abordagem pela matriz de ativação.

Além disso, houve a preocupação de verificar se, após as reamostragens nos dados, os modelos conseguiriam manter suas performances para a classe majoritária “Pastando”. Pelos resultados representados nas Tabelas 3, 4 e 5, percebe-se que, mesmo após as reamostragens, todos os modelos mantiveram o nível da performance para essa classe em todos os cenários, com diferenças desprezíveis inferiores a 1%. Portanto, a inclusão de novas amostras, seja por repetição ou artificiais, nos dados de treinamento não prejudicou a generalização dos modelos para a classe majoritária.

Comparando os resultados entre as CNNs para cada cenário, percebe-se que no cenário com a base original, os resultados foram semelhantes com a diferença em torno de 1%. O destaque foi a abordagem com a matriz de ativação por ter obtido melhores resultados nos cenários com reamostragem com as diferenças em torno de 2.6% para a abordagem com CONV1D e 1.5% para a abordagem por combinação. Constata-se que a abordagem pela matriz de ativação conseguiu um desempenho superior em relação à classe minoritária “EmPe-Ruminando” com uma diferença de 2.7% para a CNN para a abordagem com CONV1D e 5.72% para a abordagem por combinação. Isso pode ter ocorrido devido a sua maior dimensionalidade que, apesar da pouca quantidade de amostras para a classe “EmPe-Ruminando”, ainda conseguiu mapear uma melhor representabilidade dessa classe por meio da presença das ativações de alguns neurônios em determinadas regiões dessa matriz.

Entretanto, olhando os cenários com reamostragem, percebe-se que as técnicas usadas, exatamente por proporem uma igualdade de padrões entre as classes e, assim, permitirem que os classificadores alcancem uma melhor modelagem das classes minoritárias, conseguiram diminuir essa diferença, fazendo com que os resultados das abordagens por combinação e CONV1D se aproximassem dos resultados obtidos da abordagem pela matriz de ativação.

6.2 CNNs X CLASSIFICADORES DE APRENDIZAGEM DE MÁQUINA

As Tabelas 7, 8 e 9 contêm, respectivamente, as medidas de performance das CNNs junto com os cinco classificadores de aprendizagem de máquina apresentados

na Seção 2.3.3 nos três cenários propostos. É fundamental destacar que todos esses classificadores de aprendizagem de máquina foram implementados com os valores padrões dos hiperparâmetros da biblioteca *scikit-learn*¹. A Tabela 6 contém a lista dos principais hiperparâmetros com os seus respectivos valores.

Tabela 6 – Lista dos principais hiperparâmetros dos classificadores

ALGORITMO	PARÂMETROS
RF	Número de árvores: 100
	Função: gini
NB	Sem hiperparâmetros
SVM	Termo de regularização (C): 1
	Kernel: rbf
	Gamma: scale
DT	Função: gini
	Estratégia de divisão: best
LDA	Otimizador: svd
	Estimador: covariância

¹ Disponível em <<https://scikit-learn.org/stable/>>

Tabela 7 – Medidas de avaliação dos modelos - base original

	Precisão (%)	Sensibilidade (%)	F1-Score (%)	Acurácia (%)
RF	85.5	83.19	84.09	88.71
NB	56.69	57.74	55.86	67.76
SVM	73.2	68.24	68.20	79.66
DT	74.18	74.41	74.28	80.9
LDA	47.64	55.4	50.17	69.12
CNN (COMB)	77.58	76.4	76.92	83.86
CNN (MA)	79.01	76.45	77.61	84.01
LENET	80.41	80.27	80.29	85.66
ALEXNET	79.04	77.8	78.26	84.17
MINIVGG	80.86	80.55	80.67	85.91
RESNET	79.41	79.91	79.46	85.2
MOBILENET	78.7	78.42	78.5	84.09

Tabela 8 – Medidas de avaliação dos modelos - base *bootstrap*

	Precisão (%)	Sensibilidade (%)	F1-Score (%)	Acurácia (%)
RF	85.21	84.16	84.56	89.02
NB	54.57	58.60	54.77	64.78
SVM	74.66	76.25	75.22	81.84
DT	74.08	74.23	74.11	80.71
LDA	54.38	59.09	55.69	65.64
CNN (COMB)	81.05	81.02	81	86.25
CNN (MA)	81.29	83.07	81.99	86.65
LENET	81.91	81.71	81.79	86.73
ALEXNET	81.37	82.55	81.92	86.82
MINIVGG	82.06	81.79	81.91	86.76
RESNET	82.84	83.43	83.11	87.62
MOBILENET	81.85	80.28	81.02	86.34

Tabela 9 – Medidas de avaliação dos modelos - base SMOTE

	Precisão (%)	Sensibilidade (%)	F1-Score (%)	Acurácia (%)
RF	84.44	84.6	84.45	88.81
NB	55.62	59.21	56.2	66.04
SVM	75.03	76.12	75.42	82.2
DT	74.13	76.36	75.13	80.92
LDA	54.13	58.87	55.5	65.34
CNN (COMB)	80.9	81.81	80.06	85.77
CNN (MA)	82.53	82.5	82.4	87.11
LENET	81.65	82.13	81.88	86.48
ALEXNET	81.50	81.61	81.44	86.52
MINIVGG	82.46	81.74	82.09	86.92
RESNET	82.62	82.46	82.44	87.18
MOBILENET	82.24	81.33	81.76	86.76

Comparando os resultados entre as CNNs, nos diferentes cenários, com foco na medida f1-score, percebe-se o aumento do desempenho de todos os modelos nos cenários com reamostragem, o que se assemelha com os resultados mostrados e comentados na Seção 6.1. Dentre tais melhorias, destaca-se as CNNs por combinação e por matriz de ativação que conseguiram um aumento acima de 4% nos cenários com reamostragem.

Olhando isoladamente para o cenário da base original, a MiniVGG foi a CNN que obteve a melhor performance, resultando em uma diferença de 3.74% em relação a CNN por combinação, a qual foi, dentre as CNNs, que obteve o pior resultado. Entretanto, essa diferença foi reduzida nos cenários com reamostragem, resultando em pouca distinção entre as CNNs com uma diferença abaixo de 2%.

Em relação aos classificadores de aprendizagem de máquina, no geral, as CNNs obtiveram as melhores performances em todos os três cenários das bases de treinamento, com exceção da RF. A RF, dentre todos os modelos, foi o classificador que obteve os melhores resultados em todos os cenários, ratificando o porquê da sua utilização em diversos contextos. No cenário com a base original, as maiores diferenças entre a RF e a CNN por combinação (pior CNN para esse cenário) ficam acima de 7%. Entretanto, se compararmos os resultados, nos cenários com reamostragem, perceberemos uma redução desses valores aproximadamente para 3.5% no cenário *bootstrap* e 4.4% no cenário com a base SMOTE.

6.3 DISCUSSÕES

Algumas tomadas de decisões durante o desenvolvimento dos experimentos desta dissertação acabaram acarretando em algumas consequências. Primeiramente, há de se ressaltar que à medida que se aumenta a quantidade de classes de comportamentos, conseqüentemente, aumenta-se a complexidade para uma boa generalização do modelo, pois alguns comportamentos podem ser bastante similares e difíceis até para o ser humano por meio da visualização *in loco*. Essas complexidades podem ser percebidas pela análise das matrizes de confusão dos modelos. Apenas para ilustrar, a Tabela 10 representa a matriz de confusão da CNN por combinação no cenário com a base original. As outras matrizes de confusão dos modelos estão disponíveis nos *notebooks* no repositório do remoto¹. Percebe-se que o comportamento “Deitado-Ruminando” é muitas vezes classificado como “Deitado-Parado”, como também acontece o inverso. Outro exemplo é a confusão dos classificadores entre os comportamentos “EmPe-Parado” e “EmPe-Ruminando”.

Tabela 10 – Matriz confusão da CNN por combinação em relação à base de teste - cenário com a base de treino original

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2214	3	15	20	33	8
Deitado-Rumi	4	635	149	13	0	13
Deitado-Parado	17	114	510	22	0	18
EmPe-Parado	33	44	49	417	1	80
Andando	51	0	0	0	440	0
EmPe-Rumi	12	18	12	114	2	175

Diante disso, fica nítido que o acréscimo de classes pela segmentação de um determinado comportamento aumentou a complexidade do problema de classificação. Essas situações também aconteceram nas matrizes de confusão dos outros modelos nos cenários com reamostragem (Apêndices B e C). Diferentemente da quantidade média de classes de comportamentos presente nos trabalhos selecionados (em torno de 3 a 4 classes simultâneas) essa dissertação envolveu seis classes de comportamentos.

Outro ponto a se destacar foi que, apesar da pouca quantidade de amostras (segunda classe minoritária), o comportamento “Andando” se destaca por apresentar bons resultados

¹ Moncattle - <<https://github.com/andssuu/moncattle>>

entre as CNNs. Entretanto, isso se deve pela boa discriminação a partir dos dados da velocidade de deslocamento provinda do GPS que, apesar da presença de poucos dados, ainda consegue representar bem tal comportamento. Além disso, é um comportamento que se diferencia bastante do restante e não possui subdivisões tais como o “Deitado” e “EmPé”. Ainda assim, de acordo com a matriz de confusão da CNN por combinação no cenário com a base original representada na Tabela 10, visualizamos que o comportamento “Andando” foi confundido, em alguns casos, com o comportamento “Pastando”. Essa situação também se repetiu com todos os classificadores em todos os cenários, o que é compreensível, pois o animal estar em pé e em movimento em ambos comportamentos.

Por fim, se analisarmos os resultados dos modelos nos diferentes cenários, conseguimos visualizar um aumento, em particular, das performances das CNNs e da SVM gerado pela reamostragem dos dados, situação que não acontece com a RF e nem com a DT. É possível entender essa correlação dos resultados entre a DT e a RF, uma vez que a RF consiste em um conjunto de DTs por meio da técnica *bagging* (combinação de classificadores). Diante disso, formula-se a seguinte pergunta: à medida que tivéssemos mais dados, esse aumento nas medidas de avaliação tenderia a se manter?. Do ponto de vista da aprendizagem profunda, esse fenômeno é compreensível, pois os algoritmos que implementam tal abordagem, em geral, requerem muitos dados para uma efetiva modelagem do problema.

7 CONCLUSÃO

Este trabalho explorou o paradigma da aprendizagem profunda para a classificação do comportamento bovino por meio de dados de sensores inerciais embarcados em uma coleira. Após uma revisão na literatura, percebeu-se que tal paradigma ainda não é tão explorado para esse problema de classificação e que a maioria das soluções se concentram na utilização de algoritmos de aprendizagem de máquina. Os sensores inerciais utilizados foram o acelerômetro, giroscópio, magnetômetro e GPS. Seis classes de comportamentos simultaneamente foram levadas em consideração: “Pastando”, “Deitado-Ruminando”, “Deitado-Parado”, “EmPe-Parado”, “Andando” e “EmPe-Ruminando”. Devido ao desbalanceamento dos dados, houve a aplicação de técnicas de reamostragem, as quais foram pouco exploradas pelos trabalhos selecionados, para realizar uma superamostragem. As técnicas de reamostragens escolhidas para os experimentos foram o *bootstrap* e o SMOTE.

Duas abordagens para a expansão do vetor de características foram proposta baseadas nas transformações por combinação e pela matriz de ativação devida a baixa dimensão do vetor de características originário. Em relação às CNNs, a abordagem por matriz de ativação conseguiu convergir mais rápido do que a CNN por combinação, além do que, obteve melhores resultados para a classe minoritária “EmPe-Ruminando”.

Houve análises e comparações entre os resultados das CNNs e com os principais algoritmos de aprendizagem de máquina utilizados na literatura para classificação do comportamento de bovinos por meio das seguintes medidas de avaliação: acurácia, precisão, sensibilidade e f1-score.

Após as análises, conclui-se que, no geral, as reamostragens melhoraram o desempenho dos modelos, especialmente, nos casos das CNNs e SVM em relação às classes minoritárias. Além disso, não houve nenhuma redução no desempenho para as classes majoritárias. As CNNs obtiveram bons resultados e demonstraram ser uma opção a ser considerada na construção de um sistema de monitoramento de bovinos. A RF foi o classificador com melhor desempenho em todos os cenários dos experimentos. Entretanto, ao contrário das CNNs e SVM, não apresentou melhorias significativas nos cenários com as reamostragens.

Os modelos propostos possuem bastante relevância nos planejamentos táticos e

estratégicos dos criadores de bovinos, pois a identificação dos comportamentos dos mesmos são indicadores da saúde e bem-estar dos animais, além do que, podem servir também na identificação do cio das vacas e ajudar na maximização da taxa de reprodução por inseminação artificial. Portanto, todas essas informações disponíveis ao produtor em tempo real irão ajudá-lo bastante na tomada de decisão.

7.1 CONTRIBUIÇÕES

Este trabalho se propôs a apresentar uma solução computacional para um problema real no manejo de bovinos, assim, o mesmo se posiciona na fronteira de duas áreas: Computação e Ciências Agrárias. Diante disso, por si só, é um desafio, pois ambas as áreas possuem métodos específicos que precisam ser dominados por quem deseja atuar nesta fronteira. Como as principais contribuições desta dissertação, destacam-se:

- Mostrar os benefícios das técnicas de reamostragem para a classificação do comportamento de bovinos, especificamente por se tratar de um problema recorrente das bases de dados nesse contexto;
- Mostrar o potencial da abordagem da aprendizagem profunda como uma alternativa para o problema de classificação do comportamento de bovinos;
- Apresentar duas abordagens para expansão do vetor de características que podem ser utilizadas em outros problemas com baixa dimensão em outros contextos;
- Os modelos propostos podem ser parte de um sistema maior de monitoramento de bovinos e agregar valor para diversas aplicações específicas, tais como a detecção de cio das vacas, para a otimização da reprodução e conseqüentemente o aumento da produção do leite, sistemas de alarmes para o monitoramento do bem-estar e a identificação precoce de doenças, como, por exemplo, a claudicação.
- Artigos publicados, como (SANTOS et al., 2021) e (BEZERRA et al., 2021), a partir de interações com outros grupos de pesquisa que possibilitou uma melhor compreensão do contexto agrícola.

7.2 TRABALHOS FUTUROS

Em relação aos trabalhos futuros, a análise da performance para outras arquiteturas de CNNs mais robustas e conseqüentemente com maior poder de generalização é uma opção a ser tomada. Para isso, algumas heurísticas podem ser criadas com base nas técnicas *Grid Search* (PONTES et al., 2016) e *Random Search* (BERGSTRA; BENGIO, 2012).

Além disso, entendemos que outras técnicas de aprendizagem profunda também poderiam ser abordadas. Em especial, destacamos a *Recurrent Neural Network* (RNN), a qual consegue levar em consideração a ordem do estado dos comportamentos, o que poderia ser uma estratégia pertinente para a análise do comportamento de bovinos. Entretanto, para o uso efetivo dessa rede, a construção da base de dados deve seguir um processo de rotulação mais completo e rígido, pois a sequência temporal dos comportamentos precisa ser mantida, o que acaba se tornando mais um desafio além da construção do próprio modelo.

Nessa mesma perspectiva, a construção de novas bases de dados, a fim de ratificar os modelos e abordagens propostas nesta dissertação, é um caminho muito útil a se tomar. Somado a isso, outros sensores poderiam ser abordados para esse problema, como dados de vídeo ou áudio, de forma que possibilitem alternativas de soluções menos invasivas ao animal.

Em especial, a obtenção dessas novas bases de dados, com os mesmos dados dos sensores da base utilizada por este trabalho, permitiria a investigação do uso de *transfer learning* por meio das técnicas de aprendizagem profunda. Uma vez que há poucas bases de dados disponíveis contendo informações sobre o comportamento de bovinos, essa proposta agregaria valor na construção de novos modelos por meio do conhecimento extraído e armazenado nas redes executadas em diversas bases de dados.

REFERÊNCIAS

- AGRICULTURA, P. e. A. M. Ministério da. **Agropecuária é único setor da economia com crescimento na pandemia, diz IBGE**. 2021. <<https://www.gov.br/agricultura/pt-br/assuntos/noticias/agropecuaria-e-unico-setor-com-crescimento-na-pandemia-diz-ibge>>. [Online; acessado 18-Novembro-2021].
- ALOM, M. Z.; TAHA, T. M.; YAKOPCIC, C.; WESTBERG, S.; SIDIKE, P.; NASRIN, M. S.; ESESN, B. C. V.; AWWAL, A. A. S.; ASARI, V. K. The history began from alexnet: A comprehensive survey on deep learning approaches. **arXiv preprint arXiv:1803.01164**, 2018.
- ANDRIAMANDROSO, A. L. H.; LEBEAU, F.; BECKERS, Y.; FROIDMONT, E.; DUFRASNE, I.; HEINESCH, B.; DUMORTIER, P.; BLANCHY, G.; BLAISE, Y.; BINDELLE, J. Development of an open-source algorithm based on inertial measurement units (imu) of a smartphone to detect cattle grass intake and ruminating behaviors. **Computers and electronics in agriculture**, Elsevier, v. 139, p. 126–137, 2017.
- BANK, D.; KOENIGSTEIN, N.; GIRYES, R. Autoencoders. **arXiv preprint arXiv:2003.05991**, 2020.
- BATISTA, P. H. D.; ALMEIDA, G. L. P. d.; SARMENTO, R. M.; PANDORFI, H.; MELO, A. A. S. d.; ROLIM, M. M.; MEDEIROS, V. W. C. d.; GONÇALVES, G. E. Monitoramento da atividade de bovinos em pastoreio por dispositivo eletrônico de sensoriamento baseado em gps. **Revista de Ciências Agrárias**, Sociedade de Ciências Agrárias de Portugal, v. 42, n. 2, p. 31–40, 2019.
- BECCIOLINI, V.; PONZETTA, M. P. et al. Inferring behaviour of grazing livestock: Opportunities from gps telemetry and activity sensors applied to animal husbandry. **Eng. Rural Dev**, v. 17, p. 192–198, 2018.
- BENAISSA, S.; TUYTTENS, F. A.; PLETS, D.; CATTRYSSE, H.; MARTENS, L.; VANDAELE, L.; JOSEPH, W.; SONCK, B. Classification of ingestive-related cow behaviours using rumiwatch halter and neck-mounted accelerometers. **Applied Animal Behaviour Science**, Elsevier, v. 211, p. 9–16, 2019.
- BENAISSA, S.; TUYTTENS, F. A.; PLETS, D.; PESSEMIER, T. D.; TROGH, J.; TANGHE, E.; MARTENS, L.; VANDAELE, L.; NUFFEL, A. V.; JOSEPH, W. et al. On the use of on-cow accelerometers for the classification of behaviours in dairy barns. **Research in veterinary science**, Elsevier, v. 125, p. 425–433, 2019.
- BERCKMANS, D. Precision livestock farming technologies for welfare management in intensive livestock systems. **Rev. Sci. Tech**, v. 33, n. 1, p. 189–196, 2014.
- _____. General introduction to precision livestock farming. **Animal Frontiers**, Oxford University Press, v. 7, n. 1, p. 6–11, 2017.
- BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. **Journal of machine learning research**, v. 13, n. 2, 2012.

- BERNARDI, A. d. C.; JUNIOR, A. L.; PEREZ, N.; INAMASU, R. Potencial de uso das tecnologias de agricultura e pecuária de precisão e automação. **Embrapa Pecuária Sudeste-Documentos (INFOTECA-E)**, São Carlos, SP: Embrapa Pecuária Sudeste, 2017., 2017.
- BEZERRA, I. do N. B.; BATISTA, P. H. D.; SANTOS, A.; GUISELINI, C.; ALMEIDA, G. L. P. de; GONÇALVES, G.; MEDEIROS, V. Determinação do comportamento de bovinos de leite a pasto baseado em dados de localização gps e ambientais. In: **SBAI 2021 - Especiais** (). [s.n.], 2021. Disponível em: <<http://XXXXX/217092.pdf>>.
- BIRCH, P.; RAHIMI, N.; OVERBURRY, P.; YOUNG, R.; CHATWIN, C. Implementations and optimisations of optical conv2d networks designs. In: SPIE. **Semiconductor Lasers and Laser Dynamics IX**. [S.l.], 2020. v. 11356, p. 178–185.
- BISHOP, C. M. **Pattern Recognition and Machine Learning (Information Science and Statistics)**. Berlin, Heidelberg: Springer-Verlag, 2006. ISBN 0387310738.
- BISHOP-HURLEY, G.; HENRY, D.; SMITH, D.; DUTTA, R.; HILLS, J.; RAWNSLEY, R.; HELLICAR, A.; TIMMS, G.; MORSHED, A.; RAHMAN, A. et al. An investigation of cow feeding behavior using motion sensors. In: IEEE. **2014 IEEE International Instrumentation and Measurement Technology Conference (I2MTC) Proceedings**. [S.l.], 2014. p. 1285–1290.
- BURGIN, M. **Theory of information: fundamentality, diversity and unification**. [S.l.]: World Scientific, 2010. v. 1.
- CHAWLA, N. V.; BOWYER, K. W.; HALL, L. O.; KEGELMEYER, W. P. Smote: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine learning**, Springer, v. 20, n. 3, p. 273–297, 1995.
- DECANDIA, M.; GIOVANETTI, V.; ACCIARO, M.; MAMELI, M.; MOLLE, G.; CABIDDU, A.; MANCA, C.; COSSU, R.; SERRA, M.; RASSU, S. et al. Monitoring grazing behaviour of sarda cattle using an accelerometer device. **Grassland resources for extensive farming systems in marginal lands: major drivers and future scenarios**, p. 143, 2017.
- DEVALAL, S.; KARTHIKEYAN, A. Lora technology-an overview. In: IEEE. **2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)**. [S.l.], 2018. p. 284–290.
- DUDA, R. O.; HART, P. E.; STORK, D. G. **Pattern Classification**. 2. ed. New York: Wiley, 2001. ISBN 978-0-471-05669-0.
- DUTTA, R.; SMITH, D.; RAWNSLEY, R.; BISHOP-HURLEY, G.; HILLS, J.; TIMMS, G.; HENRY, D. Dynamic cattle behavioural classification using supervised ensemble classifiers. **Computers and electronics in agriculture**, Elsevier, v. 111, p. 18–28, 2015.
- EFRON, B.; TIBSHIRANI, R. J. **An introduction to the bootstrap**. [S.l.]: CRC press, 1994.

FORNEY, G. D. The viterbi algorithm. **Proceedings of the IEEE**, Ieee, v. 61, n. 3, p. 268–278, 1973.

GRUNKEMEIER, G. L.; WU, Y. Bootstrap resampling methods: something for nothing? **The Annals of thoracic surgery**, Elsevier, v. 77, n. 4, p. 1142–1144, 2004.

HAMÄLÄINEN, W.; JÄRVINEN, M.; MARTISKAINEN, P.; MONONEN, J. Jerk-based feature extraction for robust activity recognition from acceleration data. In: IEEE. **2011 11th International Conference on Intelligent Systems Design and Applications**. [S.l.], 2011. p. 831–836.

HAMILTON, A. W.; DAVISON, C.; TACHTATZIS, C.; ANDONOVIC, I.; MICHIE, C.; FERGUSON, H. J.; SOMERVILLE, L.; JONSSON, N. N. Identification of the rumination in cattle using support vector machines with motion-sensitive bolus sensors. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 19, n. 5, p. 1165, 2019.

HAN, H.; WANG, W.-Y.; MAO, B.-H. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In: SPRINGER. **International conference on intelligent computing**. [S.l.], 2005. p. 878–887.

HAN, J.; MORAGA, C. The influence of the sigmoid function parameters on the speed of backpropagation learning. In: SPRINGER. **International workshop on artificial neural networks**. [S.l.], 1995. p. 195–201.

HE, H.; BAI, Y.; GARCIA, E. A.; LI, S. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In: IEEE. **2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)**. [S.l.], 2008. p. 1322–1328.

HE, K.; ZHANG, X.; REN, S.; SUN, J. Deep residual learning for image recognition. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 770–778.

HINTON, G. E.; SRIVASTAVA, N.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. R. Improving neural networks by preventing co-adaptation of feature detectors. **arXiv preprint arXiv:1207.0580**, 2012.

HOWARD, A. G.; ZHU, M.; CHEN, B.; KALENICHENKO, D.; WANG, W.; WEYAND, T.; ANDREETTO, M.; ADAM, H. Mobilenets: Efficient convolutional neural networks for mobile vision applications. **arXiv preprint arXiv:1704.04861**, 2017.

ISMAIL, A.; AHMAD, S. A.; SOH, A. C.; HASSAN, K.; HARITH, H. H. Improving convolutional neural network (cnn) architecture (minivggnet) with batch normalization and learning rate decay factor for image classification. **International Journal of Integrated Engineering**, v. 11, n. 4, 2019.

JESUS, L. d. Identificação do comportamento bovino por meio do monitoramento animal. 2014.

JUNG, D.; LEE, J.; PARK, H. Feature expansion of single dimensional time series data for machine learning classification. In: IEEE. **2021 Twelfth International Conference on Ubiquitous and Future Networks (ICUFN)**. [S.l.], 2021. p. 96–98.

JURDAK, R.; SOMMER, P.; KUSY, B.; KOTTEGE, N.; CROSSMAN, C.; MCKEOWN, A.; WESTCOTT, D. Camazotz: multimodal activity-based gps sampling. In: **Proceedings of the 12th international conference on information processing in sensor networks**. [S.l.: s.n.], 2013. p. 67–78.

KALMAN, B. L.; KWASNY, S. C. Why tanh: choosing a sigmoidal function. In: IEEE. **[Proceedings 1992] IJCNN International Joint Conference on Neural Networks**. [S.l.], 1992. v. 4, p. 578–581.

KAMILARIS, A.; PRENAFETA-BOLDÚ, F. X. Deep learning in agriculture: A survey. **Computers and electronics in agriculture**, Elsevier, v. 147, p. 70–90, 2018.

KASFI, K. T.; HELLICAR, A.; RAHMAN, A. Convolutional neural network for time series cattle behaviour classification. In: **Proceedings of the Workshop on Time Series Analytics and Applications**. [S.l.: s.n.], 2016. p. 8–12.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. **nature**, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

LECUN, Y.; BOTTOU, L.; BENGIO, Y.; HAFFNER, P. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, Ieee, v. 86, n. 11, p. 2278–2324, 1998.

LECUN, Y. et al. Lenet-5, convolutional neural networks. **URL: <http://yann.lecun.com/exdb/lenet>**, v. 20, n. 5, p. 14, 2015.

LI, Z.; LIU, F.; YANG, W.; PENG, S.; ZHOU, J. A survey of convolutional neural networks: analysis, applications, and prospects. **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, 2021.

LIN, S.; CAI, L.; LIN, X.; JI, R. Masked face detection via a modified lenet. **Neurocomputing**, Elsevier, v. 218, p. 197–202, 2016.

LINKER, R. **Smart Collar Tag for Cow Market Forecast to 2027 - COVID-19 Impact and Global Analysis By Product Type ; Application, and Geography**. 2021. <https://www.reportlinker.com/p06038358/Smart-Collar-Tag-for-Cow-Market-Forecast-to-COVID-19-Impact-and-Global-Analysis-By-Product-Type-Application-and-Geography.html?utm_source=GNW>. [Online; accessed 2-November-2021].

LOMBA, L. F. D. Identificação do comportamento bovino a partir dos dados de movimentação e do posicionamento do animal. 2015.

LU, Y.; YOUNG, S. A survey of public datasets for computer vision tasks in precision agriculture. **Computers and Electronics in Agriculture**, Elsevier, v. 178, p. 105760, 2020.

MAATEN, L. Van der; HINTON, G. Visualizing data using t-sne. **Journal of machine learning research**, v. 9, n. 11, 2008.

MAĆKIEWICZ, A.; RATAJCZAK, W. Principal components analysis (pca). **Computers & Geosciences**, Elsevier, v. 19, n. 3, p. 303–342, 1993.

- NISBETT, W. H.; KAVURI, A.; DAS, M. On the correlation between second order texture features and human observer detection performance in digital images. **Scientific Reports**, Nature Publishing Group, v. 10, n. 1, p. 1–14, 2020.
- NORRING, M.; HÄGGMAN, J.; SIMOJOKI, H.; TAMMINEN, P.; WINCKLER, C.; PASTELL, M. Lameness impairs feeding behavior of dairy cows. **Journal of Dairy Science**, Elsevier, v. 97, n. 7, p. 4317–4321, 2014.
- PAL, M. Random forest classifier for remote sensing classification. **International journal of remote sensing**, Taylor & Francis, v. 26, n. 1, p. 217–222, 2005.
- PHI, K. P. C.; THI, K. N.; DINH, C. N.; DUC, N. T.; DUC, T. T. Classification of cow's behaviors based on 3-dof accelerations from cow's movements. **International Journal of Electrical and Computer Engineering**, IAES, v. 9, n. 3, p. 1656–1662, 2019.
- PHILLIPS, C. **Cattle behaviour and welfare**. [S.l.]: John Wiley & Sons, 2008.
- PONTES, F. J.; AMORIM, G.; BALESTRASSI, P. P.; PAIVA, A.; FERREIRA, J. R. Design of experiments and focused grid search for neural network parameter optimization. **Neurocomputing**, Elsevier, v. 186, p. 22–34, 2016.
- QUINLAN, J. R. Induction of decision trees. **Machine learning**, Springer, v. 1, n. 1, p. 81–106, 1986.
- RAHMAN, A.; SMITH, D.; HILLS, J.; BISHOP-HURLEY, G.; HENRY, D.; RAWNSLEY, R. A comparison of autoencoder and statistical features for cattle behaviour classification. In: IEEE. **2016 international joint conference on neural networks (IJCNN)**. [S.l.], 2016. p. 2954–2960.
- RAHMAN, A.; SMITH, D.; LITTLE, B.; INGHAM, A.; GREENWOOD, P.; BISHOP-HURLEY, G. Cattle behaviour classification from collar, halter, and ear tag sensors. **Information processing in agriculture**, Elsevier, v. 5, n. 1, p. 124–133, 2018.
- RAHMAN, M.; BHUIYAN, M.; KAMAL, M.; SHAMSUDDIN, M. Prevalence and risk factors of mastitis in dairy cows. **Bangladesh Veterinarian**, v. 26, n. 2, p. 54–60, 2009.
- RESEARCH; MARKETS. **Global Pedometer Market 2017-2021**. 2017. <<https://www.researchandmarkets.com/reports/4339882/global-pedometer-market-2017-2021>>. [Online; accessed 2-November-2021].
- RIABOFF, L.; AUBIN, S.; BEDERE, N.; COUVREUR, S.; MADOUASSE, A.; GOURMAND, E.; CHAUVIN, A.; PLANTIER, G. Considering pre-processing of accelerometer signal recorded with sensor fixed on dairy cows is a way to improve the classification of behaviours. In: **9. European Conference on Precision Livestock Farming (ECPLF)**. [S.l.: s.n.], 2019. p. np.
- RIABOFF, L.; AUBIN, S.; BEDERE, N.; COUVREUR, S.; MADOUASSE, A.; GOURMAND, E.; CHAUVIN, A.; PLANTIER, G. Evaluation of pre-processing methods for the prediction of cattle behaviour from accelerometer data. **Computers and Electronics in Agriculture**, Elsevier, v. 165, p. 104961, 2019.

RIABOFF, L.; POGGI, S.; MADOUASSE, A.; COUVREUR, S.; AUBIN, S.; BÉDÈRE, N.; GOUMAND, E.; CHAUVIN, A.; PLANTIER, G. Development of a methodological framework for a robust prediction of the main behaviours of dairy cows using a combination of machine learning algorithms on accelerometer data. **Computers and Electronics in Agriculture**, v. 169, p. 105179, 2020.

RISH, I. et al. An empirical study of the naive bayes classifier. In: **IJCAI 2001 workshop on empirical methods in artificial intelligence**. [S.l.: s.n.], 2001. v. 3, n. 22, p. 41–46.

RODRIGUEZ-BAENA, D. S.; GOMEZ-VELA, F. A.; GARCÍA-TORRES, M.; DIVINA, F.; BARRANCO, C. D.; DAZ-DIAZ, N.; JIMENEZ, M.; MONTALVO, G. Identifying livestock behavior patterns based on accelerometer dataset. **Journal of Computational Science**, Elsevier, v. 41, p. 101076, 2020.

SANTOS, A. da S.; SILVA, J. V. da; MEDEIROS, V. W. C. de; GONÇALVES, G. E. Previsão e construção de indicadores operacionais da produção de cana-de-açúcar por meio de séries temporais. In: SBC. **Anais do XII Workshop de Computação Aplicada à Gestão do Meio Ambiente e Recursos Naturais**. [S.l.], 2021. p. 117–126.

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

SHAHRIAR, M. S.; SMITH, D.; RAHMAN, A.; FREEMAN, M.; HILLS, J.; RAWNSLEY, R.; HENRY, D.; BISHOP-HURLEY, G. Detecting heat events in dairy cows using accelerometers and unsupervised learning. **Computers and electronics in agriculture**, Elsevier, v. 128, p. 20–26, 2016.

SIFRE, L.; MALLAT, S. Rigid-motion scattering for texture classification. **arXiv preprint arXiv:1403.1687**, 2014.

SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, 2014.

SMITH, D.; DUTTA, R.; HELLICAR, A.; BISHOP-HURLEY, G.; RAWNSLEY, R.; HENRY, D.; HILLS, J.; TIMMS, G. Bag of class posteriors, a new multivariate time series classifier applied to animal behaviour identification. **Expert Systems with Applications**, Elsevier, v. 42, n. 7, p. 3774–3784, 2015.

SMITH, D.; LITTLE, B.; GREENWOOD, P. I.; VALENCIA, P.; RAHMAN, A.; INGHAM, A.; BISHOP-HURLEY, G.; SHAHRIAR, M. S.; HELLICAR, A. A study of sensor derived features in cattle behaviour classification models. In: IEEE. **2015 IEEE SENSORS**. [S.l.], 2015. p. 1–4.

TANG, Y.; ZHANG, Y.-Q.; CHAWLA, N. V.; KRASSER, S. Svms modeling for highly imbalanced classification. **IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)**, IEEE, v. 39, n. 1, p. 281–288, 2008.

UNGAR, E. D.; RUTTER, S. M. Classifying cattle jaw movements: comparing iger behaviour recorder and acoustic techniques. **Applied animal behaviour science**, Elsevier, v. 98, n. 1-2, p. 11–27, 2006.

- WANG, X.; TANG, X. Random sampling lda for face recognition. In: IEEE. **Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.** [S.l.], 2004. v. 2, p. II–II.
- WEI, G.; LI, G.; ZHAO, J.; HE, A. Development of a lenet-5 gas identification cnn structure for electronic noses. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 19, n. 1, p. 217, 2019.
- WILLIAMS, L.; JACKSON, E.; BISHOP-HURLEY, G.; SWAIN, D. Drinking frequency effects on the performance of cattle: A systematic review. **Journal of animal physiology and animal nutrition**, Wiley Online Library, v. 101, n. 6, p. 1076–1092, 2017.
- WILLIAMS, L. R.; BISHOP-HURLEY, G. J.; ANDERSON, A. E.; SWAIN, D. L. Application of accelerometers to record drinking behaviour of beef cattle. **Animal Production Science**, CSIRO, v. 59, n. 1, p. 122–132, 2019.
- WILLIAMS, M. L.; JAMES, W. P.; ROSE, M. T. Variable segmentation and ensemble classifiers for predicting dairy cow behaviour. **biosystems engineering**, Elsevier, v. 178, p. 156–167, 2019.
- WINOGRAD, S. On computing the discrete fourier transform. **Mathematics of computation**, v. 32, n. 141, p. 175–199, 1978.
- WORLD, R. R. **GLOBAL ACCELEROMETERS MARKET REPORT, HISTORY AND FORECAST 2016-2027, BREAKDOWN DATA BY MANUFACTURERS, KEY REGIONS, TYPES AND APPLICATION.** 2021. <<https://www.researchreportsworld.com/global-accelerometers-market-18331839>>. [Online; accessed 2-November-2021].
- YUNTA, C.; GUASCH, I.; BACH, A. Lying behavior of lactating dairy cows is influenced by lameness especially around feeding time. **Journal of dairy science**, Elsevier, v. 95, n. 11, p. 6546–6549, 2012.
- ZHANG, H. The optimality of naive bayes. **AA**, v. 1, n. 2, p. 3, 2004.
- ZHAO, Z. A.; LIU, H. **Spectral feature selection for data mining.** [S.l.]: Taylor & Francis, 2012.
- ZHONG, L.; HU, L.; ZHOU, H. Deep learning based multi-temporal crop classification. **Remote sensing of environment**, Elsevier, v. 221, p. 430–443, 2019.
- ZHU, Y.; LI, G.; WANG, R.; TANG, S.; SU, H.; CAO, K. Intelligent fault diagnosis of hydraulic piston pump combining improved lenet-5 and pso hyperparameter optimization. **Applied Acoustics**, Elsevier, v. 183, p. 108336, 2021.
- ZHUANG, F.; QI, Z.; DUAN, K.; XI, D.; ZHU, Y.; ZHU, H.; XIONG, H.; HE, Q. A comprehensive survey on transfer learning. **Proceedings of the IEEE, IEEE**, v. 109, n. 1, p. 43–76, 2020.

APÊNDICES

APÊNDICE A – Arquiteturas das CNNs após otimização

Tabela A.1 – Arquitetura CNN por combinação

Camada	Dimensão da saída	Tamanho do filtro	Stride
Entrada	10x10x1		
Conv2D (Relu)	10x10x32	9x9x32	1
Batch Normalization	10x10x32		
MaxPooling2D	5x5x32	2x2	2
Dropout	5x5x32		
Conv2D (Relu)	5x5x96	3x3	1
Batch Normalization	5x5x96		
MaxPooling2D	2x2x96	2x2	2
Dropout	2x2x96		
Flatten	1x384		
Dense (Relu)	1x384		
Batch Normalization	1x384		
Dropout	1x384		
Dense (Relu)	1x128		
Batch Normalization	1x128		
Dropout	1x128		
Dense (Softmax)	1x6		

Tabela A.2 – Arquitetura CNN por matriz da ativações

Camada	Dimensão da saída	Tamanho do filtro	Stride
Entrada	1x10		
Flatten	1x10x1		
Dense (Relu)	1x1024x1		
Batch Normalization	1x1024x1		
Reshape	32x32x1		
Conv2D (Relu)	32x32x32	9x9x32	1
Batch Normalization	32x32x32		
MaxPooling2D	16x16x32	2x2	2
Dropout	16x16x32		
Conv2D (Relu)	16x16x96	3x3	1
Batch Normalization	16x16x96		
MaxPooling2D	8x8x96	2x2	2
Dropout	8x8x96		
Flatten	1x6144		
Dense (Relu)	1x384		
Batch Normalization	1x384		
Dropout	1x384		
Dense (Relu)	1x128		
Batch Normalization	1x128		
Dropout	1x128		
Dense (Softmax)	1x6		

APÊNDICE B – Matrizes de confusão da CNN com abordagem por combinação

Tabela B.1 – Matriz confusão CNN (combinação) em relação à base de teste - cenário com base de treino original

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2214	3	15	20	33	8
Deitado-Rumi	4	635	149	13	0	13
Deitado-Parado	17	114	510	22	0	18
EmPe-Parado	33	44	49	417	1	80
Andando	51	0	0	0	440	0
EmPe-Rumi	12	18	12	114	2	175

Deitado-Rumi: classe “Deitado-ruminando”.

EmPe-Rumi: classe “EmPe-Ruminando”.

Tabela B.2 – Matriz confusão CNN (combinação) em relação à base de teste - cenário com base de treino bootstrap

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2202	2	7	19	57	6
Deitado-Rumi	7	666	106	26	1	8
Deitado-Parado	19	109	514	31	0	8
EmPe-Parado	35	30	36	450	2	71
Andando	26	0	0	2	463	0
EmPe-Rumi	10	17	5	77	3	221

Deitado-Rumi: classe “Deitado-ruminando”.

EmPe-Rumi: classe “EmPe-Ruminando”.

Tabela B.3 – Matriz confusão CNN (combinação) em relação à base de teste - cenário com base de treino smote

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2152	4	16	66	49	6
Deitado-Rumi	4	709	56	30	0	15
Deitado-Parado	12	148	469	44	0	8
EmPe-Parado	25	26	23	469	1	80
Andando	36	0	0	7	448	0
EmPe-Rumi	5	14	2	66	2	244

Deitado-Rumi: classe “Deitado-ruminando”.

EmPe-Rumi: classe “EmPe-Ruminando”.

APÊNDICE C – Matrizes de confusão da CNN com abordagem por matriz de ativação

Tabela C.1 – Matriz confusão CNN (matriz de ativações) em relação à base de teste - cenário com base de treino original

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2243	3	3	18	25	1
Deitado-Rumi	8	613	160	25	0	8
Deitado-Parado	31	112	489	43	0	6
EmPe-Parado	52	35	21	439	2	75
Andando	63	0	0	1	427	0
EmPe-Rumi	10	18	2	113	2	188

Deitado-Rumi: classe “Deitado-ruminando”.

EmPe-Rumi: classe “EmPe-Ruminando”.

Tabela C.2 – Matriz confusão CNN (matriz de ativações) em relação à base de teste - cenário com base de treino bootstrap

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2193	4	8	25	58	5
Deitado-Rumi	4	697	81	13	0	19
Deitado-Parado	16	94	527	30	1	13
EmPe-Parado	39	31	25	425	4	100
Andando	26	0	2	0	462	1
EmPe-Rumi	13	5	3	43	2	267

Deitado-Rumi: classe “Deitado-ruminando”.

EmPe-Rumi: classe “EmPe-Ruminando”.

Tabela C.3 – Matriz confusão CNN (matriz de ativações) em relação à base de teste - cenário com base de treino smote

	Pastando	Deitado-Rumi	Deitado-Parado	EmPe-Parado	Andando	EmPe-Rumi
Pastando	2209	3	9	29	37	6
Deitado-Rumi	4	671	88	23	0	28
Deitado-Parado	16	73	548	32	0	12
EmPe-Parado	36	15	29	456	1	87
Andando	59	1	2	6	420	3
EmPe-Rumi	8	13	1	53	1	257

Deitado-Rumi: classe “Deitado-ruminando”.

EmPe-Rumi: classe “EmPe-Ruminando”.