



Universidade Federal Rural da Amazônia
Pró-Reitoria de Pesquisa e Desenvolvimento Tecnológico
Programa de Pós-Graduação em Saúde e Produção Animal na Amazônia

MANUELLA PAULA DE MESQUITA NUNES

DESENVOLVIMENTO DE UM MÉTODO DE CLASSIFICAÇÃO DA
CONFORMAÇÃO DE CARCAÇAS DE OVINOS BASEADA NO *K-
NEAREST NEIGHBOR* (KNN)

BELÉM

2017

MANUELLA PAULA DE MESQUITA NUNES

DESENVOLVIMENTO DE UM MÉTODO DE CLASSIFICAÇÃO DA
CONFORMAÇÃO DE CARCAÇAS DE OVINOS BASEADA NO *K-
NEAREST NEIGHBOR* (KNN)

Dissertação apresentada à Pró-Reitoria de Pesquisa e Desenvolvimento Tecnológico e ao Programa de Pós-Graduação em Saúde e Produção Animal na Amazônia, como pré-requisito para obtenção do título de Mestre em Saúde e Produção Animal na Universidade Federal Rural Da Amazônia.

Orientador:

D.Sc. André Guimarães Maciel e Silva

Co-orientadores:

D.Sc Emerson Cordeiro Morais

DSc. Luciana Celi da Silva Chaves

BELÉM

2017

Nunes, Manuella Paula de Mesquita
Desenvolvimento de um método de classificação da
conformação de carcaças de ovinos baseada no *k-nearest neighbor*
(knn) / Manuella Paula de Mesquita Nunes . – Belém, PA, 2017.
71 f.

Dissertação (Mestrado em Saúde e Produção Animal na Amazônia) –
Universidade Federal Rural da Amazônia, 2017.

Orientador: Prof. Dr. André Guimarães Maciel e Silva; Co-
orientadores: Dr. Emerson Cordeiro Morais; Dr. Luciara Celi da
Silva Chaves.

1. Carcaça Ovina. 2. Carcaças Ovinas - Classificação e
Tipificação. 3. *k-Nearest Neighbor* (KNN) - Algoritmo. 4. Avaliação
na Carcaça. I. Silva, André Guimarães Maciel e Silva, (orient.) II.
Morais, Emerson Cordeiro (coorient.). III. Chaves, Luciara Celi da Silva
(coorient.). IV. Título

CDD – 664.92931

MANUELLA PAULA DE MESQUITA NUNES

**DESENVOLVIMENTO DE UM MÉTODO DE CLASSIFICAÇÃO
DA CONFORMAÇÃO DE CARCAÇAS DE OVINOS BASEADA NO
K-NEAREST NEIGHBOR (KNN)**

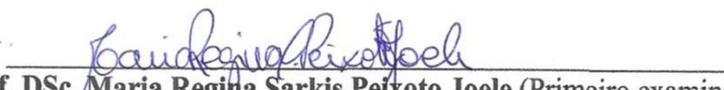
Dissertação apresentada à Pró-Reitoria de Pesquisa e Desenvolvimento Tecnológico e ao Programa de Pós-Graduação em Saúde e Produção Animal na Amazônia, como exigência para obtenção do título de Mestre em Saúde e Produção Animal na Universidade Federal Rural da Amazônia. Área de concentração: Produção Animal.

Aprovado em 31 de março de 2017.

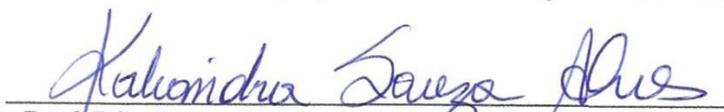
BANCA EXAMINADORA



Prof. DSc. André Guimarães Maciel e Silva (Orientador - Presidente)
Universidade Federal do Pará – UFPA



Prof. DSc. Maria Regina Sarkis Peixoto Joele (Primeiro examinador)
Instituto Federal de Educação, Ciência e Tecnologia do Pará – IFPA



Prof. DSc. Kaliandra Souza Alves (Segundo examinador)
Universidade Federal Rural da Amazônia – UFRA



Prof. DSc. Shirley Motta de Souza (Terceiro examinador)
Instituto Federal de Educação, Ciência e Tecnologia do Pará – IFPA

BELÉM

2017

“O segredo da saúde mental e corporal está em não se lamentar pelo passado, não se preocupar com o futuro, nem se adiantar aos problemas, mas viver sabiamente e seriamente o presente.”
_Buda

AGRADECIMENTOS

Nós só vamos entender por que tudo tem um propósito, quando certas coisas acontecem na nossa vida. Já diria um autor desconhecido, em suas sábias palavras, que mesmo quando tudo parece desabar, cabe a nós decidirmos entre rir ou chorar, ir ou ficar, desistir ou lutar, porque no caminho incerto da vida, mais importante é o decidir.

Agradeço,

À Deus,

Aos meus pais Manoel e Edilza Nunes, pela paciência, pelas palavras de motivação, de que sempre as coisas darão certo e que cada coisa há seu tempo. Que em momentos de dificuldades pelo qual passei nesse período de mestrado, tive um apoio incondicional. A minha irmã Rafaella pela paciência e pelas conversas tarde da noite.

Aos meus familiares, exemplo de união, amor, diversidade que sempre estarão lá dividindo amor e acolhendo sempre quem precise dando força e alegria.

Aos meus amigos desde calouros, Andréa, Victor, pela amizade até mesmo depois da graduação, cada um com seu apoio distinto, mesmo tomando caminhos diferentes agradeço a presença.

Ao Professor André Silva, orientador, pela oportunidade de entrada em seu grupo de pesquisa, o oferecimento do local de experimente/ trabalho de campo, pela oportunidade de elaboração deste trabalho, me proporcionando melhores oportunidades acadêmicas, pela paciência e compreensão.

A Professora Luciara Chaves, que mais uma vez se mostrou, além de uma orientadora e mãe de pesquisa, uma amiga. Pela dedicação mesmo em tempos que era somente dela. . Pelos conselhos, paciência, apoio, incentivo e dedicação durante esse período. Pela oportunidade de elaboração e acompanhamento deste trabalho, me proporcionando uma nova visão de vida acadêmica e profissional

Ao Professor Emerson Morais, co-orientador, pelos puxões de orelha, pelos ensinamentos e rigorosidade de formatação, e pela oportunidade de desenvolver um estudo inovador como esse, expandido meu conhecimento em outras áreas que a Zootecnia pode oferecer. E ser a primeira pessoa a me dizer “Bem-vinda a Pós-Graduação”

A equipe PROSEG, por se apresentar minha segunda família, mais uma vez. Aos que já fizeram parte e aos contemporâneos, pelo convívio e pela ajuda, sempre disponíveis nas horas difíceis e de correria. Por fazer os momentos de trabalho serem mais agradáveis, divertidos e hilários, apesar do cansaço de todos. Em especial, aos meus amigos e auxiliares nesse trabalho, Gabriel Xavier e Gabriel Duarte que me ajudaram a seguir em frente e a finalizar esse trabalho, em meus momentos de dificuldade, agradeço de todo coração.

Ao PROSEG como grupo de pesquisa, agradeço por disponibilizar um ambiente de pesquisa em que me construí e fui construída profissionalmente durante 6 anos. Sempre serei grata.

Aos Doutorandos Marco Antonio e Alyne e ao Mestrando Jonas pelo convívio diário durante o experimento, pelas conversas, puxões de orelhas, ensinamentos de vida e tiradas de sarro, que amenizaram o cansaço de um experimento.

Aos Professores e Funcionários ligados ao IFPA (Instituto Federal de Educação, Ciência e Tecnologia do Pará/Castanhal-PA) e do Instituto de Medicina Veterinária da UFPA (Universidade Federal do Pará/Castanhal-PA) que contribuíram para que esse trabalho fosse realizado.

Aos professores do PPGSPA (Programa de Pós-Graduação em Saúde e Produção Animal) e aqueles advindos de outros locais, que contribuíram para minha formação.

Aos membros que compuseram a banca avaliadora, agradeço a participação, correções, sugestões e críticas que contribuíram para o engrandecimento dessa dissertação.

A CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) por tornar possível minha manutenção durante a pós graduação, por meio de bolsa de mestrado.

Aos meus colegas também mestrandos, aos funcionários do setor da pós-graduação, e todos aqueles que contribuíram direta ou indiretamente para a conclusão deste trabalho.

RESUMO

A seleção de animais voltados para a produção de carne tem como objetivo a visualização do conjunto de características responsáveis pelo aumento da eficiência econômica da exploração com vistas a atender às expectativas do consumidor. Seguindo este princípio, o valor de uma determinada carcaça, isto é, a sua aceitabilidade frente a uma série de exigências do mercado, depende da quantidade e da qualidade da porção comestível. A avaliação da conformação, que expressa os perfis musculares, é uma ferramenta utilizada para classificar carcaças quanto ao seu valor comercial. Nesse contexto, o emprego de ferramentas computacionais de processamento de dados em busca de padrões pode otimizar esse tipo de avaliação. Com isso, objetiva-se, neste estudo, aplicar uma metodologia para classificar carcaças ovinas quanto à sua conformação a partir de atributos como peso, medidas morfométricas e de área do olho-de-lombo (AOL) de carcaças utilizando o classificador inteligente *k-Nearest Neighbor* (KNN). Foram avaliadas 72 carcaças de cordeiros castrados mantidos sob sistema de confinamento pertencentes à rebanhos comerciais localizados no estado do Pará. Um total de 29 medidas morfométricas foram tomadas nas carcaças frias inteiras, além do peso e da AOL. Estas foram classificadas de acordo com o peso em leves e pesadas e, posteriormente, segundo sua conformação adotando-se o sistema SEUROP por um avaliador utilizando de imagens digitais da carcaça fria. A execução do algoritmo KNN seguiu a metodologia de validação cruzada *leave-one-out*, onde utilizaram-se diferentes estratégias de aprendizado para prever a conformação das carcaças, cujos conjuntos de treinamento se distinguiam em número de exemplos (carcaças), de atributos (variáveis) Nos conjuntos numéricos obteve-se PECC de 61,11% (k=7) para a estratégia 1; 62,50% (k=5) para a estratégia 2; 50,94% (k=5 e 23) para a estratégia 3 e 58,49% (k=11) para a estratégia 4. Nos conjuntos nominais referentes às Estratégias 1, 2, 3 e 4 (k=1) observou-se PECC de 100%. O algoritmo KNN, foi capaz de prever os escores de conformação a partir de medidas morfométricas. A partir dos resultados da acurácia, da sensibilidade das classes e da precisão da classificação, utilizando o conjunto numérico, a melhor estratégia foi a estratégia 2 (CONF_72_CP), com redução de atributos, conseguiu otimizar a predição de classificação em relação as demais estratégias.

Palavras-Chave: Carcaças. Conformação. Medidas morfométricas. Predição.

ABSTRACT

The selection of meat production animals has the objective of visualizing the set of characteristics responsible for increasing the economic efficiency of the farm in order to meet consumer expectations. Following this principle, the value of a given carcass, that is, its acceptability to a series of market requirements, depends on the quantity of the quality of the edible portion. The evaluation of the conformation, which expresses the muscle profiles, is a tool used to classify carcasses in terms of commercial value. In this context, the use of computational tools for data processing in search of patterns can optimize this type of evaluation. The objective of this study is to apply a methodology that has the capacity to classify ovine carcasses in terms of their conformation based on attributes such as weight, morphometric and area of the loin-eye area (LEA) of carcasses using the *k-Nearest Neighbor (KNN)* smart classifier. A total of 72 carcasses of castrated lambs kept under a feedlot system belonging to commercial herds located in the state of Pará were evaluated. A total of 29 morphometric measurements were taken on whole cold carcasses, in addition to weight and LEA. These were classified according to the weight in light and heavy and, later, according to their conformation adopting the SEUROP system by an evaluator by means of digital images of the cold carcass. The execution of the *KNN* algorithm followed the leave-one-out cross validation methodology, where different learning strategies were used to predict the conformation of the carcasses, whose training sets were distinguished in number of examples (carcasses), attributes (variables). In the numerical sets, PECC of 61.11% ($k = 7$) was obtained for strategy 1; 62.50% ($k = 5$) for strategy 2; 50.94% ($k = 5$ and 23) for strategy 3 and 58.49% ($k = 11$) for strategy 4. In the nominal sets referring to Strategies 1, 2, 3 and 4 ($k = 1$) %. The *KNN* algorithm was able to predict the conformation scores from morphometric measurements. From the results of accuracy, class sensitivity and classification accuracy, the best strategy was strategy 2 (CONF_72_CP), with attribute reduction, able to optimize the prediction of classification in relation to the other strategies.

Keywords: Carcasses. Conformation. Morphometric measurements. Prediction.

LISTA DE FIGURAS

Figura 1 – Classificação de carcaças ovinas de esquema de carcaça pesada pelo sistema de classificação por conformação S.E.U.R.O.P.	20
Figura 2 – Diferenças entre os limites de decisão dos algoritmos NN e 5NN para um conjunto de exemplos com 2 dimensões ou atributos e 3 classes diferentes.....	26
Figura 3– Esquema representando a metodologia de amostragem aleatória estratificada para uma relação de 70:30 entre treinamento e teste.....	28
Figura 4 – Esquema da metodologia de validação cruzada <i>leave-one-out</i>	33
Figura 5 – Medidas Morfométricas tomadas nos painéis Frontal e Lateral.	38
Figura 6 - Interface do ImageJ com imagem digitalizada da área do olho-de-lombo em papel vegetal mostrando a configuração da escala, feita por meio de régua graduada em centímetros utilizada como referencia para definir a distância conhecida (<i>Known distance</i>) por meio da marcação de uma reta (em amarelo).....	39
Figura 7 – Resultado da mensuração da área do olho-de-lombo, contornada em amarelo, onde o resultado é visualizado na janela <i>Results</i> na coluna <i>Area</i>	39
Figura 8 – Interface do NetBeans IDE 8.1 exibindo as guias Projetos, onde encontram-se os arquivos necessários à execução do algoritmo, as abas referentes às classes do programa e a guia Saída.....	40
Figura 9 – Ilustração do processo de formação dos bancos de dados.	42
Figura 10– Janela do NetBeans IDE 8.1 mostrando o arquivo CONFO_CP_72.data que contém todos os 72 exemplos (carcaças), um por linha, cujos atributos (variáveis) e a classe (escore de conformação) são nominais com valores separados somente por vírgula.....	42
Figura 11- Janela do NetBeans IDE 8.1 mostrando o arquivo CONFO_CP_72.names que contém a descrição dos atributos (variáveis) e da classe (escore de conformação) utilizados.....	43
Figura 12 – Interface do <i>software</i> WEKA versão 3.8.0, para o processamento do conjunto numérico.	44
Figura 13 – Distribuição e porcentagem relativa ao total por classe de carcaças ovinas do banco de dados de carcaças leves e pesadas (n=72) e somente com carcaças pesadas (n=53).	49

LISTA DE TABELAS

Tabela 1 – Conformação ou perfis de carcaças ovinas segundo o Regulamento (Comunidade Europeia) nº. 22/2008.	20
Tabela 2 – Demonstração de uma matriz de confusão de um classificador.	31
Tabela 3 – Descrição das medidas morfométricas mensuradas na carcaça de ovinos de acordo com os painéis Frontal e Lateral.	37
Tabela 4 – Descrição das classes de categorização dos atributos (medidas morfométricas) provenientes da carcaça.	41
Tabela 5 – Exemplo de Estrutura das características dos bancos de dados de Conformação formam o arquivo de exemplos “.data”.	42
Tabela 6 – Exemplo de Estrutura das características dos bancos de dados que formam o arquivo de atributos “.names”.	43
Tabela 7 – Descrição estatística dos atributos utilizados e coeficiente de correlação de cada atributo com o escore de conformação.	50
Tabela 8 - Medidas morfométricas com maior autovetor dentro de cada componente principal selecionado.	52
Tabela 9 – Acurácia do algoritmo KNN (PECC) utilizando a metodologia de validação <i>cross validation leave-one-out</i> (LOO) para a classificação do escore de conformação de carcaças ovinas com diferentes classes de dados, dimensionalidades e número de exemplos.	55
Tabela 10 - Comparação entre os fatores esquema de carcaça, quantidade de atributos e classe de dados na acurácia ou acurácia da classificação do escore de conformação de carcaças ovinas pelo teste de Wilcoxon e normalidade dos dados pelo teste de Shapiro-Wilk.	56
Tabela 11 - Matrizes de confusão para as estratégias de classificação do escore de conformação de carcaças ovinas processadas com conjuntos numéricos (CONF_72_NUM), por k de maior acurácia na estratégia.	57
Tabela 12 - Matrizes de confusão para as estratégias de classificação do escore de conformação de carcaças ovinas processadas com conjuntos nominais (CONF_72_NOM), por k de maior acurácia na estratégia.	58

SUMÁRIO

1	INTRODUÇÃO	13
2	ESTUDOS RELACIONADOS À CARCAÇA OVINA	16
2.1	A Carcaça Ovina.....	16
2.2	Classificação e tipificação de carcaças ovinas	17
2.2.1	Conformação	18
2.3	Avaliações na carcaça.....	21
2.3.1	Morfometria.....	21
2.3.2	Área de olho-de-lombo.....	22
3	RECONHECIMENTO DE PADRÕES.....	23
3.1	A Zootecnia de Precisão e o Reconhecimento de Padrões.....	23
3.2	Conjuntos, exemplos e atributos.....	24
3.3	<i>k-Nearest Neighbor (KNN)</i>	25
3.3.1	Princípio do algoritmo KNN	26
3.3.2	Medidas de similaridade entre exemplos.....	27
3.3.3	Particularidades do algoritmo KNN	27
3.4	Esquema geral para classificação com o algoritmo KNN	29
3.5	Métodos de avaliação da predição.....	30
3.5.1	Avaliação quantitativa	30
3.5.2	Avaliação preditiva.....	32
3.6	Análise de componentes principais (ACP).....	33
4	MATERIAL E MÉTODOS	36
5	RESULTADOS E DISCUSSÃO	49
6	CONCLUSÃO.....	60
7	REFERÊNCIAS BIBLIOGRÁFICAS	61
	APÊNDICE	69

1 INTRODUÇÃO

A criação de ovinos para corte no Brasil é desenvolvida utilizando diversos sistemas de criação, desde sistemas intensivos com alta tecnologia em nutrição, genética e ambiência, até sistemas extensivos quase que em estado selvagem (SILVA, 2010). O Nordeste possui a liderança no panorama nacional com expansão contínua dos rebanhos. No Norte, os rebanhos despontaram em número de cabeças a partir dos anos 2000 (RAINERI, 2014).

Visando esse crescimento, a capacidade de atender às expectativas do consumidor fornecendo produtos de qualidade e manter a consistência dos produtos é a base para o sucesso em um mercado altamente competitivo (CHANDRARATNE et al., 2007). Desta forma, a seleção de animais voltados para corte tem como objetivo a visualização do conjunto de características que devam ser melhoradas, responsáveis pelo aumento da eficiência econômica da exploração (SOUSA, 2008).

Essa seleção visa a obtenção de carcaças com máximo desenvolvimento muscular, agregando valor já que esse tecido constitui maior proporção da porção comestível e de valor econômico (SANTOS e PÉREZ, 2000). Uma carcaça ideal possui perfis convexos, deposição moderada de gordura (OSÓRIO, 1992), visando sua aceitabilidade frente às exigências do mercado consumidor.

Considerando que os sistemas de criação e as exigências do mercado consumidor apresentam constantes mudanças, para Suguisawa (2002), a busca por tamanho e peso ideal do animal para produção de carne é contínua.

Logo, para fins de definir o valor de uma determinada carcaça, isto é, a sua aceitabilidade frente a uma série de exigências do mercado, a tipificação representa uma ferramenta bem mais apurada (SAINZ e ARAÚJO, 2001).

Desta forma uma correta avaliação da carcaça é imprescindível para o estabelecimento de um sistema de classificação e tipificação de carcaça, atendendo ao mercado consumidor de carne cada vez mais exigente e crescente em questões de demanda (CEZAR E SOUSA, 2007; CLEMENTE, 2012).

O monitoramento de características que estimem a conformação e o acabamento da carcaça auxilia, assim, não somente na escolha dos animais a serem abatidos e fornece informações úteis para serem incorporadas tanto em modelos de crescimento

quanto de seleção animal (BERGEN *et al.*, 1996), como também zela pela qualidade do produto que será disponibilizado ao mercado.

O aumento da produtividade animal está diretamente relacionado a um volume maior de informações, o que torna difícil e complexo o crescimento da atividade sem nenhum auxílio direto na organização dessas respostas (SILVA, 2007). Assim, o emprego de ferramentas computacionais de controle e monitoramento de problemáticas zootécnicas se torna uma realidade no setor produtivo, como por exemplo, a análise de sinais e redes neurais para estudos de comportamento e bem-estar, e rastreabilidade de produtos de origem animal.

Segundo Bandeira Filho (2003), a Zootecnia de Precisão é direcionada para a tomada de decisões e para ações mais precisas do que aquelas anteriormente tomadas, que se baseavam apenas em “valores médios” ou “valores típicos”, ou seja, atua principalmente na busca por ferramentas avançadas que garantam um melhor aproveitamento produtivo, controle, monitoramento da produção animal, e a gestão da qualidade do produto, possibilitando desta forma agregar valor ao produto final.

Ricardo (2011), o principal propósito de um sistema de classificação de carcaças é de facilitar a comercialização pela descrição das características de carcaça mais importantes comercialmente.

Dessa forma, a classificação da conformação de carcaças ovinas é realizada sob características subjetivas, onde o avaliador humano diferencia aquelas que se apresentam mais compactas e de perfis convexos, notadamente mais bem conformadas, daquelas mais compridas e de perfis mais côncavos, menos conformadas. Esses perfis musculares podem ser expressos por meio de medidas morfométricas e biométricas, comprovadamente preditoras da quantidade de músculo na carcaça (ARAUJO, 2014; McMANUS *et al.* 2013; PINHEIRO e JORGE, 2010).

Os algoritmos de aprendizado de máquina, como o *KNN*, são capazes de identificar, em um grande volume de informações, padrões que sintetizam o conhecimento humano e a partir deles gerar teorias ou modelos que sejam úteis na solução de problemas futuros (WITTEN, 2011).

Sendo assim, a aplicação dessas ferramentas de mineração de dados na classificação automática de carcaças ovinas pode não só otimizar a avaliação de animais

com aptidão para a produção de carne como também extrair informações relevantes para a ciência animal sobre a relação entre um modelo subjetivo, como o sistema SEUROP, e um modelo de objetivo, baseado na combinação multivariada de medidas.

Portanto, no cenário competitivo atual, construir e gerenciar conhecimento de apoio a especialistas no controle de processos de maneira geral pode ser útil para uma unidade de produção, seja industrial seja agropecuária, principalmente quando se trata de processos que ocorrem sob incertezas e com dados incompletos (SILVA, 2007).

Objetiva-se, neste estudo, propor uma metodologia de classificação do escore de conformação de carcaças ovinas por meio de informações concatenadas de características quantitativas de carcaça e medidas morfométricas utilizando o algoritmo *k-Nearest Neighbor (KNN)* ou k-vizinhos mais próximos.

2 ESTUDOS RELACIONADOS À CARÇAÇA OVINA

2.1 A Carçaça Ovina

A tendência da indústria alimentar é fornecer aos mercados produtos de alta qualidade para satisfazer as demandas dos consumidores, sendo que a padronização é dos principais objetivos para a obtenção de um produto de qualidade (GOYACHE et al., 2001).

Segundo a Portaria nº 307 de dezembro de 1990, define-se como carçaça caprina e ovina, o corpo inteiro do animal abatido, sangrado, esfolado, eviscerado, desprovido de cabeça, patas, glândulas mamárias, verga, exceto suas raízes e testículos, retiram-se os rins e as gorduras perirrenais e inguinais, por fim no rabo, permanecem não mais do que seis vértebras coccígeas.

No sistema de produção de carne, a carçaça é o elemento mais importante do animal, porque nela está contida a porção comestível de maior valor comercial (PIRES et al., 1999). Para Viana (2008), por esse motivo, a comparação de características facilita a identificação de diferenças existentes entre as carçaças, apontando os diferentes padrões existentes. E isso, de certa forma, beneficia os setores de comercialização, pois há um avanço na busca na seleção de animais que apresentem carçaças com boa deposição de tecidos comestíveis.

De acordo com Chandraratne *et al.* (2007), a avaliação visual é o principal componente de categorização do sistema de classificação e tipificação de carçaças, porém esse tipo de avaliação é considerada subjetiva e altamente variável, não sendo capaz, em sua maioria, de detectar atributos comercialmente importantes de carçaças. Então, o desenvolvimento de um preciso, confiável, e robusto sistema de controle seria uma maneira de melhorar a abordagem de classificação visual atual.

Portanto, técnicas de mensurações, como pesos vivos e avaliações subjetivas conformação, tem melhorado a seleção de características de carçaça (LAMBE et al., 2008). Sendo assim, o estudo da conformação da carçaça é um componente importante nas negociações entre os produtores de carne os operadores do mercado (DÍEZ et al., 2006).

2.2 Classificação e tipificação de carcaças ovinas

A classificação de carcaças consiste em juntar as carcaças semelhantes em classes segundo critérios ou características vinculadas, normalmente, aos animais abatidos que lhes deram origem, tais como espécie, sexo, idade/maturidade e peso, por exemplo. Os sistemas de classificação são de natureza cardinal, onde as classes são identificadas por nomes, números ou símbolos neutros que não tem nenhum significado de hierarquização para as carcaças, ou seja, a classificação não estabelece relações de superioridade ou inferioridade, em termos quantitativos ou qualitativos, entre as diferentes classes de carcaças, sendo mais um processo de identificação das carcaças (CEZAR e SOUSA, 2010).

A importância da classificação e tipificação da carcaça são importantes para organizar e facilitar o sistema de compra e venda (comercialização do produto). Também auxilia na padronização destes, visto que através da classificação e da tipificação pode-se definir o valor de determinada carcaça, forçando toda a cadeia produtiva a adaptar-se para produzir o padrão que irá resultar em melhor remuneração, o que conseqüentemente irá aumentar a qualidade e o rendimento da carne produzida no país (BRIDI, 2002).

Portanto, de acordo com Bridi (2002), a tipificação e classificação podem ser usadas para várias finalidades como: auxiliar na comercialização entre produtores – frigoríficos – varejo, garantir ao consumidor especificações diferenciadas de cortes e produtos, auxiliar a indústria frigorífica quanto ao destino dado à carcaça, e exportação, mercado interno, fabricação de produtos “light”, venda “*in natura*”.

Entre os métodos de avaliação de carcaças, o SEUROP é o mais consolidado, possuindo a legislação em vigor referente à classificação de carcaças de ovino regida pelo Regulamento (Comunidade Europeia) N° 22/2008, consiste na avaliação da conformação e no revestimento de gordura subcutânea avaliados de maneira visual subjetiva, exigindo profissionais capacitados para sua realização.

Segundo esse método, as carcaças são classificadas inicialmente em duas categorias: carcaças de ovinos com menos de 12 meses de idade (L), quando estes não trocam sua dentição de leite pela permanente e outros ovinos (S), animais com mais de 12 meses de idade. Posteriormente são subdivididas em esquema de carcaça, sendo que quando apresenta o peso de carcaça superior a 13,0 kg, é denominada Esquema de

carcaça pesada (ECP), sendo classificadas quanto a sua conformação pelo desenvolvimento dos perfis da carcaça, nomeadamente das suas principais partes (perna, dorso e paleta), pela quantidade de tecido adiposo no exterior da carcaça e no interior da cavidade torácica. Para peso de carcaças com peso inferior a 13 kg, é dada a denominação Esquema de carcaças Leves (ECL), utilizando peso de carcaça, cor da carne e camada de gordura como critérios de classificação.

A tipificação consiste em diferenciar as carcaças em tipos hierarquizados ou ordenados de acordo com aspectos quantitativos e qualitativos da carne presente na carcaça. A tipificação quantitativa se baseia em características que são indicadoras de produção (rendimento) de carne pela carcaça, enquanto a tipificação qualitativa se fundamenta em caracteres preditores da palatabilidade da carne originada por cada tipo de carcaça.

Para tanto, a tipificação quantitativa utiliza principalmente a conformação (musculosidade) e o acabamento (gordura externa ou de cobertura) para predizer a quantidade da porção comestível da carcaça, ou seja, estimar a carnosidade (relação músculo+gordura/osso) da carcaça, enquanto que a tipificação qualitativa usa o marmoreio, cor, textura e o pH para indicar a qualidade da carne (CEZAR e SOUSA, 2010).

Em geral, os sistemas de tipificação são de natureza ordinal, onde os tipos são identificados por nomes, símbolos ou números que denotam ordenação ou hierarquização das carcaças, ou seja, a tipificação estabelece relações de superioridade ou inferioridade, em termos de quantidade e qualidade, entre os diferentes tipos de carcaças (CEZAR e SOUSA, 2010).

2.2.1 Conformação

A conformação expressa o desenvolvimento das massas musculares, sendo um parâmetro obtido pela verificação do perfil muscular, os quais definem anatomicamente as regiões de uma carcaça (MCMANUS *et al.*, 2013; SAINZ, 2000).

Define-se como conformação a forma que a carcaça adquire como resultado da quantidade e distribuição da massa muscular e gordura sobre a base óssea, descrita subjetivamente por meio de perfis ou contornos externos e, objetivamente através de

medições lineares e circulares, na qual as proporções das medidas dependem das relações teciduais existentes na carcaça (CEZAR e SOUSA, 2010).

Para Gonzaga Neto (2005), a conformação é um critério de qualidade que tem sido adotada nos sistemas de avaliação de carcaças e, em cada sistema sua importância está dependente da relação com o preço e qualidade, pode ser definida como a espessura de sua massa muscular (musculosidade da carcaça) em relação ao tamanho do esqueleto que os suportam.

Uma carcaça bem conformada indica alta proporção músculo:osso, apresentando em cada região anatômica uma dominância dos perfis convexos (deposição de músculo muito boa), sobretudo no traseiro, e dos diâmetros transversais frente aos perfis côncavos (deposição de músculos ruim) e dos diâmetros longitudinais, ou seja, com uma forma curta, larga, redonda e compacta. Ao contrário, as carcaças longas e estreitas são aquelas consideradas de conformação deficiente (OSÓRIO e OSÓRIO, 2004). Portanto, o desenvolvimento muscular se expressa na morfologia da carcaça, que pode ser mensurada por meio de medidas lineares que caracterizam determinada classe de conformação.

As carcaças dos pequenos ruminantes apresentam perfis que variam de côncavo até híper-convexo (excelente deposição de músculos), em função da profundidade da massa muscular depositada sobre a base óssea, que é o esqueleto. Os ovinos lanados (Texel, por exemplo), apresentam musculatura dupla e que, por conseguinte, são os únicos que geram carcaças com perfil híper-convexo, já as raças de ovinos deslanados (Dorper, Santa Inês e outras) podem produzir até carcaças convexas (CEZAR e SOUSA, 2010).

O sistema SEUROP avalia a conformação de acordo com o desenvolvimento dos perfis de quarto traseiro, dorso e quarto dianteiro da carcaça, conforme a Tabela 1 e Figura 1.

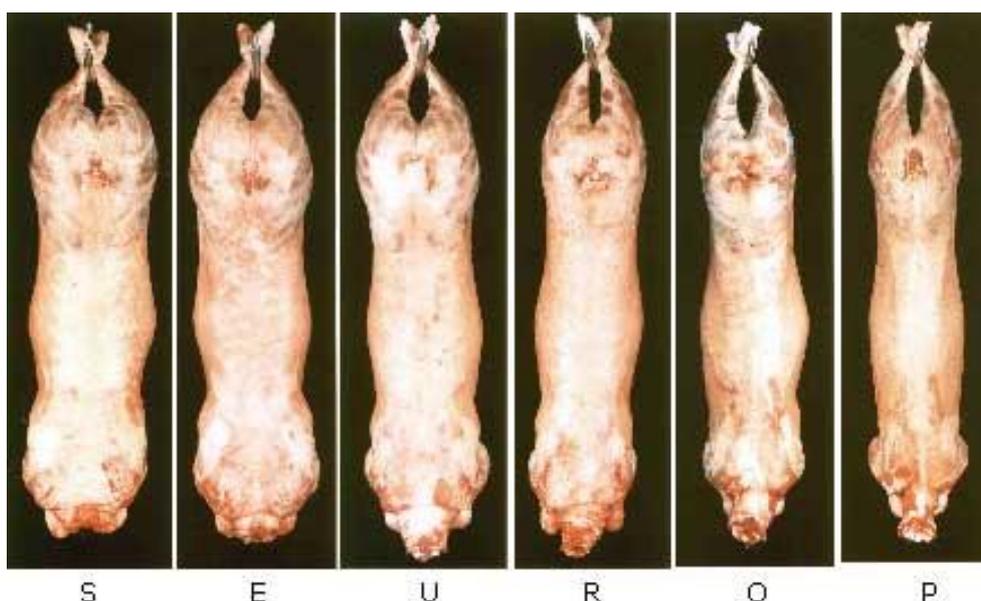
Ricardo *et al* (2012), avaliando 252 carcaças ovinas até doze meses de idade, em estudo de classificação de carcaças ovinas pelo modelo EUROP, não encontrou carcaças com o escore S, e apenas 1,4% classificadas com o escore E. Segundo o estudo, levando em consideração o peso vivo ao abate (PVA) as carcaças que apresentaram a classe de conformação U (muito boa) foram provenientes de animais

com PVA superior (44,23 kg) as demais classes, seguido pelas carcaças com classe R, com PVA de 36,39 kg. As menores classes de conformação, O (razoável) e P (mediocre), foram atribuídas às carcaças de animais com PVA inferior (31,30 e 30,58 kg).

Tabela 1 – Conformação ou perfis de carcaças ovinas segundo o Regulamento (Comunidade Europeia) n.º 22/2008.

Classes de Conformação	Descrição
S – Superior (Superior)	Todos os perfis extremamente convexos; desenvolvimento muscular excepcional com duplos músculos.
E – Excelente (Excellent)	Todos os perfis convexos a superconvexos; desenvolvimento muscular excepcional.
U – Muito Boa (Very Good)	Perfis em geral convexos; forte desenvolvimento muscular.
R – Boa (Good)	Perfil em geral retilíneo; bom desenvolvimento muscular.
O – Média (Ordinary)	Perfil retilíneos a côncavos; desenvolvimento muscular médio.
P – Médiocre (Poor)	Todos os perfis côncavos a muito côncavos; reduzido desenvolvimento muscular.

Figura 1 – Classificação de carcaças ovinas de esquema de carcaça pesada pelo sistema de classificação por conformação S.E.U.R.O.P.



2.3 Avaliações na carcaça

2.3.1 Morfometria

O estudo das carcaças é uma avaliação de parâmetros relacionados com medidas objetivas e subjetivas em relação à mesma e deve estar ligado aos aspectos e atributos inerentes à porção comestível (CARVALHO e PÉREZ, 2001).

As medidas morfométricas estão entre os parâmetros fundamentais para a avaliação objetiva da carcaça ovina (YAMAMOTO, 2006). Por meio das medidas de comprimento, largura, espessura e profundidade, expressam o dimensionamento da carcaça e possibilita a avaliação objetiva da conformação.

De acordo com Esteves (2011), carcaças que apresentam formas convexas ou arredondadas indicam maior desenvolvimento de massas musculares, enquanto que aquelas de aspecto côncavo denotam baixo desenvolvimento de tecido muscular.

Nesse sentido, a partir da realização dessas medidas, pode-se determinar uma estimativa das características da carcaça (SILVA e PIRES, 2000) o que demonstra ser essencial para estipular as partes de maior valor comercial contendo considerável condição muscular, dado que este tecido é a maior parte da porção comestível de uma carcaça (CORNÉLIO, 2014).

Em estudos com carcaças ovinas avaliando o efeito do genótipo (três genótipos diferentes, Santa Inês e seu cruzamento com os genótipos Ille de France e Texel) e do sexo e características quantitativas da carcaça Cardoso *et al.* (2013) concluíram que os genótipos Santa Inês x Texel apresentaram a maior comprimento de carcaça em relação ao Santa Inês, assim como os machos apresentaram a maior média para a mesma variável em relação às fêmeas.

Os índices zootécnicos que levam em consideração os tipos raciais, idade ao abate, sistemas de nutrição, entre outros, tem melhor acurácia quando são aplicadas medidas morfométricas, já que elas permitem a comparação entre esses parâmetros e ainda as correlações com outras medidas e com os tecidos que constituem a carcaça (CORNELIO, 2014). Bezerra et al. (2012) concluíram, em experimento com cordeiros não castrados cuja dieta continha diferentes níveis de inclusão de palma forrageira, que o perímetro da perna foi maior para carcaças mais pesadas, obtidas para 50% de inclusão.

Em estudo, Ricardo (2011), identificou que carcaças de ovinos com menos de 12 meses de idade com diferentes escores de conformação apresentaram dimensões também distintas o comprimento externo da carcaça e o perímetro da perna foram maiores para as classes bem mais conformadas esta ultima medida bastante relacionada com o desenvolvimento muscular.

2.3.2 Área de olho-de-lombo

Uma característica importante e que pode ser medida de maneira objetiva é a Área de olho-de-lombo (AOL) por predizer a quantidade de massa muscular que constitui a carcaça, além de que esse tecido está relacionado com a parte comestível e interessa para a venda no mercado. (ZUNDT *et al.*, 2003)

Este parâmetro pode ser determinado na meia carcaça por meio de um corte transversal entre a 12^a e a 13^a a costela, expondo a secção transversal do músculo *Longissimus dorsi* e obtendo-se as medidas: largura máxima do músculo, profundidade máxima do músculo e, menor e maior espessura de gordura sobre o *Longissimus dorsi* (MARQUES *et al.* 2007).

Pesquisadores envolvidos em estudos relacionados à condição corporal da carcaça obtiveram resultados que indicam a área de olho-de-lombo é uma região representativa da composição corporal do animal e pode ser utilizada na comparação entre os diversos tipos de manejo e diferentes grupos genéticos (LOUVANDINI *et al.*, 2006).

Para Jesus (2007) a área de olho-de-lombo tem correlação positiva com a quantidade aproveitável da carne para venda. De acordo com Gonzaga Neto *et al.* (2006), a AOL é uma medida que indica a quantidade de carne comercializável, portanto, a avaliação dessa área, e com EGS juntamente com outras características medidas no animal vivo, tais como o peso e altura da cernelha, podem auxiliar na estimativa da composição corporal dos animais, conseqüentemente, na estimativa do rendimento de carcaça ao abate (ROUSE *et al.*, 2000).

3 RECONHECIMENTO DE PADRÕES

3.1 A Zootecnia de Precisão e o Reconhecimento de Padrões

O conceito de Zootecnia de Precisão está relacionado, segundo Alencar Nääs (2001), à redução otimizada das perdas, bem como ao incremento da aplicação e gestão da qualidade do produto, dentro de todo o processo gerenciado. Seria, portanto, uma ferramenta que fornece meios ao produtor para o monitoramento de seus empreendimentos de forma prática para alcançar índices produtivos com base em informações geradas por sistemas especialistas.

A zootecnia de precisão torna-se uma ferramenta efetiva de controle mais rígido do sistema produtivo, diminuindo perdas de informações tornando a cadeia produtiva mais competitiva, empreendedora e com produtos de qualidade (PANDORFI, 2012).

Nesse contexto, novas ferramentas matemáticas têm sido utilizadas a serviço da produção animal, a mineração de dados, por exemplo, tem se mostrado como ferramenta promissora no entendimento de dados que monitor eventos com grande potencial de aleatoriedade, como determinação de estro e qualidade de ambiente (ALENCAR NÄÄS, 2001).

A mineração de dados é uma ciência que utiliza de métodos estatísticos para reconhecer padrões em bancos de dados. Em *data mining*, esses dados são armazenados continuamente e a busca é automatizada por um computador. Os dados são analisados continuamente visando a solução de problemas e a otimização de processos (WITTEN, 2011).

Já a utilização de Redes Neurais Artificiais (RNA) tem bom potencial de aplicação no reconhecimento de padrões relacionados, por exemplo, a comportamentos, vocalização, e na seleção de medições para um estudo particular, podendo prever o conjunto de comportamentos que um animal pode apresentar de forma detalhada e descritiva (ALENCAR NÄÄS, 2001).

Isto porque a rede neural tem a capacidade de generalização, ou seja, é capaz não apenas de classificar as entradas para as quais ela recebe treinamento, mas também de generalizar e classificar entradas que não tenham sido apresentadas ou incompletas (GIMENEZ, 2011). Chandraratne (2007), avaliou a capacidade de modelos de redes

neurais artificiais em prever diferentes tipos de carcaças de cordeiro utilizando banco de imagens e análise de textura de costeleta de cordeiro.

Outra técnica empregada para o reconhecimento de padrões são os algoritmos de classificação baseados no vizinho mais próximo (*k-Nearest Neighbor*). Sua base geral consiste em encontrar os k exemplos rotulados (vizinhos) mais próximos de um exemplo não classificado.

Gimenez (2001) utilizou o algoritmo *KNN* para realizar os testes de classificação, avaliando o espelho nasal bovino com o intuito de comprovar que esta é uma característica única e permanente do animal, podendo assim, ser sua identificação.

O Reconhecimento de Padrões é o campo da ciência que tem por objetivo a classificação de exemplos em um determinado número de categorias ou classes a partir da observação de suas características (THEODORIDIS e KOUTROUMBAS, 2003). Segundo Duda *et al.*(2001), esse recurso visa construir uma representação mais simples de um conjunto de dados através de suas características mais relevantes, possibilitando sua partição em classes.

O classificador é a função que calcula as medidas de similaridade entre os objetos no espaço de características. Existem inúmeros classificadores diferentes e a escolha ou projeto do mais adequado depende especificamente de cada caso. De forma geral, pode-se dividir os classificadores em dois tipos principais: os estatísticos (paramétricos e não-paramétricos) e os conexionistas (redes neurais) (SILVA, 2005).

3.2 Conjuntos, exemplos e atributos

Os conjuntos representam os bancos de dados que contém as informações das quais se deseja extrair alguma informação. Esses conjuntos podem conter números, nomes, datas, registros de uma infinidade de características que representam um objeto do mundo real que possui um conceito (WITTEN, 2011), o conhecimento que pode ser representado por modelos estatísticos capazes de prever novas situações.

Os conjuntos possuem, necessariamente, exemplos e atributos. Os exemplos ou instâncias são conjuntos de atributos. Os atributos ou variáveis podem ser entendidos como características de um exemplo que o torna diferente dos demais. Essa diferença pode ser expressa por classes que podem variar em número, de nenhuma a várias (*multilabeled instances*).

Uma distinção importante pode ser feita: quando os conjuntos não possui diferenciação, as técnicas de *data mining* utilizadas são as não supervisionadas ou *unsupervised learning*. Como exemplo, a análise de cluster onde grupos de exemplos são classificados baseando-se em regiões particulares calculadas por similaridade (WITTEN, 2011).

De outra forma, quando existem classes, distinções entre os exemplos que podem ser "ensinadas" por um professor ao aprendiz, diz-se que aprendizado é supervisionado (*supervised learning*) (BATISTA, 2003), cujos modelos de predição podem ser aperfeiçoados pelo professor.

Syed (2014) divide os conjuntos de dados em homogêneos, aqueles que apresentam dados de mesma natureza (ou nominais ou numéricos) e heterogêneos, aqueles que possuem uma variedade maior de atributos. Para cada um, são utilizadas metodologias diferentes quanto à métrica de cálculo da distância.

3.3 *k*-Nearest Neighbor (KNN)

O *k*-Nearest Neighbor ou *k*-vizinho mais próximo é um classificador de aprendizado supervisionado *lazy* que utiliza medidas de similaridade para classificar um exemplo desconhecido a partir dos exemplos de treinamento. O algoritmo foi desenvolvido pela necessidade de se realizar análises discriminantes quando as estimativas paramétricas de densidade são desconhecidas ou de difícil avaliação (WEBB, 2002).

Ao considerar que os dados estejam num mesmo espaço de dimensão d , que é igual ao número de atributos, o algoritmo KNN nos permite avaliar a distância entre os exemplos, que são pontos ou coordenadas, e os atributos são conjuntos de eixos ortogonais (GAMA, 2015).

O KNN possui a seu favor a facilidade de obtenção e implementação em diversos *softwares* e scripts disponíveis gratuitamente na internet, além de ser fácil entendimento e efetivo em conjuntos grandes e com ruído (IMANDOUST e BOLANDRAFTAR, 2013). A flexibilidade em usar diversas métricas de cálculo da similaridade também é interessante.

Alguns fatores que desfavorecem seu uso são o alto custo computacional, visto que cada exemplo de teste deve ser comparado com todos os outros do treinamento,

processo que pode ser lento em grande conjuntos. O KNN é sensível à medidas redundantes, sendo uma alternativa a redução de atributos (IMANDOUST e BOLANDRAFTAR, 2013).

3.3.1 Princípio do algoritmo KNN

Segundo Murty e Devi (2011), classificadores que empregam essa técnica usam alguns ou todos os padrões disponíveis no conjunto de treino para classificar um padrão de teste. A regra mais simples é classificar um exemplo desconhecido a partir de um único vizinho mais próximo (*k-nearest neighbour* ou NN) baseando-se no pressuposto teórico de que exemplos contidos em um mesmo espaço pertençam a uma mesma classe e o erro de classificação seja igual a zero.

Quando exemplos de classes diferentes se sobrepõem em um mesmo espaço, o NN perde sua eficiência. Dessa forma, a solução é utilizar os *k*-vizinhos mais próximos (*k-nearest neighbour* ou KNN) baseando-se na classe mais frequente ou majoritária (WANG et al., 2007).

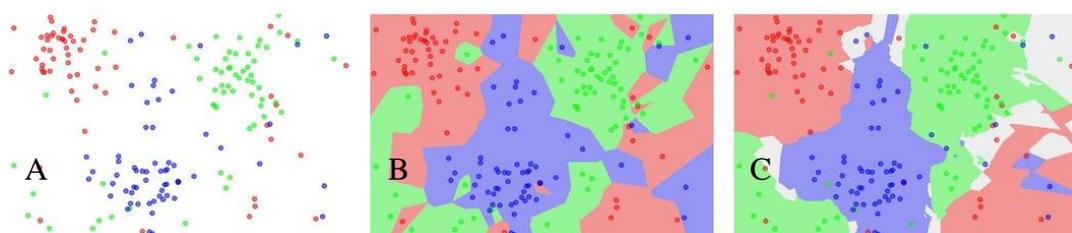
Na Figura 2, esse conceito é exemplificado em um conjunto de exemplos com 2 dimensões ou atributos e 3 classes diferentes, representadas pelas cores vermelha, verde e azul. Em A, pode-se visualizar a dispersão dos dados. Em B e C, os limites de decisão ou *decision boundaries* obtidos quando os classificadores NN e 5NN são aplicados, respectivamente.

Como descrito por Hechenbichler e Schliep (2004), um conjunto de dados de treinamento pode ser representado formalmente por

$$L = \{(y_i, x_i), i = 1, \dots, n_L\}$$

onde y_i representa as classes, x_i representa os valores dos atributos utilizados na classificação e i corresponde ao número de exemplos.

Figura 2 – Diferenças entre os limites de decisão dos algoritmos NN e 5NN para um conjunto de exemplos com 2 dimensões ou atributos e 3 classes diferentes.



A métrica de cálculo da distância entre um exemplo desconhecido (y, x) e o seus k -vizinhos mais próximos $(y_{(k)}, x_{(k)})$ é representada por $d(., .)$, onde:

$$d(x, x_{(k)}) = \min_i(d(x, x_i))$$

ou seja, a distância mínima entre os exemplos. A classe predita $\hat{y} = y_{(j)}$ é classificada com base nas y classes consideradas, decidindo pela mais frequente.

3.3.2 Medidas de similaridade entre exemplos

3.3.2.1 Distância euclidiana

A distância Euclidiana (Equação 1) normalmente calculada a partir de atributos numéricos e é a mais comum. Como pressuposto de sua utilização, recomenda-se normalizar os dados colocando-os entre 0 e 1, evitando que atributos com escalas muito diferentes possam prejudicar a classificação. A normalização para cada atributo pode ser feita pela Equação, verificando-se os valores mínimos e máximos (WITTEN, 2011).

$$d(x_i, x_k) = \left(\sum_{s=1}^p (x_{is} - x_{ks})^2 \right)^{\frac{1}{2}} \quad (\text{Eq 1})$$

$$a_i = \frac{v_i - \text{mínimo } v_i}{\text{máximo } v_i - \text{mínimo } v_i}$$

3.3.3 Particularidades do algoritmo KNN

3.3.3.1 O parâmetro k

O número de vizinhos mais próximos, identificado por k , possui influências no resultado da classificação. Quanto menor o valor de k , maior é a influência do espaço onde estão os vizinhos mais próximos, que se torna mais restrito, e a variância aumenta na medida em que o erro diminui.

Gimenez (2011) afirma que quando se utiliza apenas 1 vizinho mais próximo, a classificação é irrelevante devido à utilização de um único elemento no treinamento. De fato, isso desconsidera a grande maioria dos exemplos do treinamento. Quando se consideram mais vizinhos, diminui-se o efeito de *outliers* e de exemplos que estão se sobrepondo (*overlap*).

Normalmente, opta-se pela escolha de valores de k ímpares como critério de desempate, haja vista que, o KNN decide pela classe majoritária entre os vizinhos mais próximos. (ANCHALIA e ROY, 2014).

Em Witten (2011), é descrito que o valor de k deve ser maior caso se disponha de muitos exemplos, pois quando k e o número n de instâncias tendem ao infinito ($k/n \rightarrow 0$), a probabilidade de erros se aproxima do mínimo teórico estipulado para o conjunto.

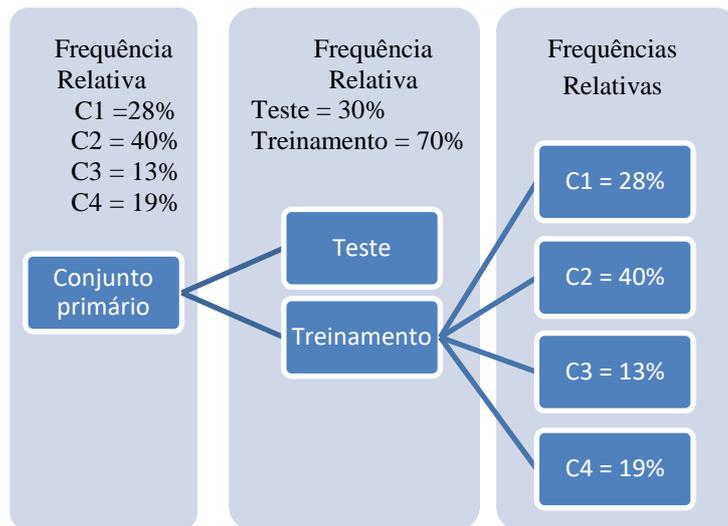
3.3.3.2 Classes desbalanceadas

Num conjunto de treinamento onde as classes são desbalanceadas ocorre uma maior frequência de uma majoritária em detrimento das demais generalizando erroneamente exemplos de classes minoritárias (RESENDE, 2005). Esses sistemas falham na predição de classes menos frequentes em detrimento das mais frequentes, que apresentam maior precisão (BATISTA, 2003).

Um treinamento ideal seria aquele em que o número de exemplos por classes é representativo (WITTEN, 2011), mas quando os exemplos são sorteados ao acaso nos conjuntos de teste e treinamento, isso pode não ocorrer.

Uma forma de minimizar o efeito nesse caso é aplicar uma amostragem aleatória: tome as frequências relativas de cada classe no conjunto primário; determine a relação treinamento:teste para dividir os exemplos, como a 70:30; sorteie os 70% dos exemplos do treinamento de forma que a frequência nesse conjunto seja igual à frequência relativa do conjunto primário (Figura 3).

Figura 3– Esquema representando a metodologia de amostragem aleatória estratificada para uma relação de 70:30 entre treinamento e teste.



3.3.3.3 Dimensionalidade do conjunto

A dimensionalidade representa a quantidade de atributos contidos num conjunto, representada por d . Geralmente, alguns atributos utilizados podem não possuir grande importância (WITTEN, 2011) onde a sua eliminação pode diminuir o custo computacional (SCHMITT, 2005) e melhorar a acurácia do classificador.

O ideal seria obter a menor quantidade de atributos possível para representar um objeto ou situação, entretanto é possível que haja perda de informação (SCHMITT, 2005). A análise de componentes principais (ACP) é uma técnica multivariada utilizada para resumir uma grande quantidade de variáveis em componentes que apresentam a maior parcela da variância total (Seção 3.4).

3.4 Esquema geral para classificação com o algoritmo KNN

Para a aplicação da metodologia de classificação utilizando o algoritmo KNN, se faz necessário seguir as seguintes etapas:

- Aquisição dos dados: atividade referente à coleta de campo. As carcaças (exemplos) pesadas, mensuradas e tipificadas, obtendo-se as variáveis (atributos) para compor o banco de dados;
- Pré-processamento: onde o banco de dados é avaliado e preparado para a classificação. Nesse tópico foram abordados como os dados foram categorizados, como ocorreu a formulação das entradas para o algoritmo e a utilização da análise de componentes principais como ferramenta de redução do número de atributos;

c) Classificação: definido pela escolha do método de amostragem, divisão da amostra, e classificação dos dados dos exemplos através do algoritmo KNN e suas diferentes estratégias;

d) Avaliação: onde o algoritmo é avaliado quanto ao seu desempenho nas diferentes estratégias utilizadas.

3.5 Métodos de avaliação da predição

O algoritmo pode prever a classe de um exemplo desconhecido de diversas maneiras em resposta à diferentes valores de k , dimensões do conjunto e exemplos de treinamento. Avaliar qual dessas predições é a mais realista para determinada situação é essencial para a otimização do classificador.

Syed (2014) explica que existem metodologias de avaliação descritiva da acurácia do classificador, como a análise de matriz de confusão e da curva ROC, que visam mais as classes, assim como é possível realizar avaliações preditivas por meio da avaliação de diferentes combinações de conjuntos de treinamento, como as metodologias mais comuns *holdout*, validação cruzada ou *cross validation* (CV) e a validação cruzada *leave-one-out* (LOOCV).

3.5.1 Avaliação quantitativa

3.5.1.1 Matriz de confusão

A matriz de confusão, segundo Syed (2014), oferece uma análise efetiva de modelos de classificação, exibindo o número de classificações corretas em relação ao número de classificações feitas pelo algoritmo. Ela pode ser construída tendo-se um número de classes igual ou maior que 2, onde as linhas indicam classificações corretas e as colunas, as preditas.

Na Tabela 3, temos descrita uma matriz de confusão, onde para cada k classe distinta (C_1, C_2, \dots, C_k), cada elemento $M(C_i, C_j)$ corresponde ao número de exemplos da base de dados que tenham sido classificados na classe C_j e que pertencem à classe C_i (PASSOS, 2005).

Em resumo temos na vertical (Verdadeira C_1, C_2, \dots, C_n , as classificações reais, e na horizontal (Predita C_1, C_2, \dots, C_n) temos as variáveis classificadas pelo algoritmo. Existe uma variedade de métricas que podem ser obtidas com a matriz de confusão. Utilizou-se somente a acurácia, que é a probabilidade de acerto do classificador.

Tabela 2 – Demonstração de uma matriz de confusão de um classificador.

Classes	Predita C ₁	Predita C ₂	Predita C ₃	Predita C ₄
Verdadeira C ₁	M(C ₁ ,C ₁)	M(C ₁ ,C ₂)	...	M(C ₁ ,C _k)
Verdadeira C ₂	M(C ₂ ,C ₁)	M(C ₂ ,C ₂)	...	M(C ₂ ,C _k)
...	M(C ₁ ,C _k)
Verdadeira C _k	M(C _k ,C ₁)	M(C _k ,C ₂)	...	M(C _k ,C _k)

A partir dessa matriz podem ser calculados a sensibilidade ou *recall*, a precisão e a porcentagem de exemplos corretamente classificados (PECC) ou acurácia.

A precisão é definida como a proporção de exemplos verdadeiros positivos (VP), ou seja, aqueles corretamente classificados, em relação à todas as classificações da classe, incluindo verdadeiros positivos e os falsos positivos (FP), aqueles erroneamente classificados em uma mesma classe. A precisão é calculada por:

$$\text{Precisão} = \frac{VP}{VP+FP}$$

A sensibilidade é proporção de verdadeiros positivos em relação aos mesmos somados aos falsos negativos (FN), exemplos de uma classe incorretamente classificados em outras classes. A sensibilidade é calculada por:

$$\text{Sensibilidade} = \frac{VP}{VP + FN}$$

A PECC é a proporção entre os exemplos corretamente classificados e todos os elementos de treinamento. Analisando-se a matriz de confusão, a PECC% é equivalente à acurácia, ou seja, é a medida de qualidade de um algoritmo classificador.

$$PECC = \frac{M(C_1, C_1) + M(C_2, C_2) + \dots + M(C_k, C_k)}{N}$$

Onde N é igual a todos os elementos de treinamento.

ou

$$PECC\% = \frac{\text{exemplos corretamente classificados}}{\text{todos os exemplos}}$$

3.5.2 Avaliação preditiva

3.5.2.1 *Houdout*

Na metodologia *houdout*, os exemplos do conjunto primário são divididos aleatoriamente em teste e treinamento, onde se preconiza que o treinamento seja maior. Dessa maneira, comumente se utiliza 1/3 para testar e 2/3 para treinar o classificador. Esse método possui maior facilidade de execução, mas a representatividade do treinamento é prejudicada.

É possível diminuir o erro da classificação executando o *houdout* aleatoriamente mais de uma vez utilizando uma relação de 1/2 para testar e 1/2 para treinar o classificador (WITTEN, 2011), o que não seria o ideal.

3.5.2.2 Validação cruzada ou *cross validation* (CV)

O *cross validation* consiste em dividir o conjunto primário em n subconjuntos (*folds*) de igual tamanho onde se utilizam $n-1$ subconjuntos a cada iteração ou execução (WITTEN, 2011). Dessa forma, se forem utilizados 10 subconjuntos, 90% dos exemplos são utilizados para o treinamento. O erro estimado é a média dos erros de cada iteração, o que é interessante para avaliar a performance do classificador e a representatividade do conjunto como um todo.

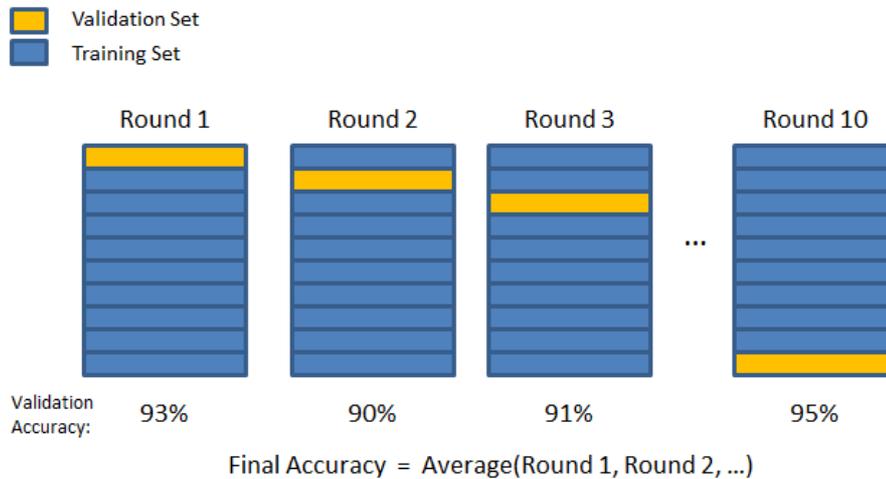
Quando um número muito grande de *folds* é utilizado, em relação ao tamanho do conjunto primário, o número de classes do conjunto de teste diminui consideravelmente, podendo ocorrer de uma classe ser alocada nesse grupo e não possuir parâmetro para classificação.

Dessa forma, quanto maior é o conjunto, menor é o número de *folds* visando maior poder de generalização. É citado que essa metodologia é útil em conjuntos com centenas ou milhares de exemplos (SYED, 2014).

3.5.2.3 Validação cruzada *leave-one-out* (LOOCV)

A Figura 4 mostra um esquema de como ocorre a divisão dos subconjuntos na metodologia LOOCV, onde cada subconjunto (*Validation set*) representa um único exemplo, sendo os outros utilizados para o treinamento (*Training set*) a cada iteração (*Round*). A LOOCV é um caso especial da validação cruzada onde o número de subconjuntos é igual ao número de exemplos do conjunto. Da mesma forma, a acurácia final (*Final accuracy*) é a média (*average*) das iterações.

Figura 4 – Esquema da metodologia de validação cruzada *leave-one-out*.



F

Fonte: mccormickml.com.

Witten (2011) cita que essa metodologia é interessante por utilizar a maior parte dos exemplos para treinamento e por ela ser determinística, ou seja, não há nenhuma casualidade. Usualmente, é utilizada em conjuntos com menos de 100 exemplos (SYED, 2014).

3.6 Análise de componentes principais (ACP)

A análise de componentes principais (ACP) é provavelmente a técnica mais antiga e mais bem conhecida dentre as técnicas de análise multivariada. Ela foi primeiramente introduzida por Pearson (1901) e desenvolvida independentemente por Hotelling (1933). Assim como muitos métodos multivariados, ela não foi amplamente usada até o advento de computadores eletrônicos, mas é agora bem consolidada em praticamente todos os pacotes estatísticos computacionais (JOLLIFFE, 1972; 1973).

A ideia central da ACP é reduzir a dimensionalidade do conjunto de dados que consiste de um grande número de variáveis interrelacionadas, mantendo assim o máximo possível da variação presente no conjunto de dados. Isso é atingido pela transformação de um novo conjunto de variáveis, os componentes principais (CP's), os quais não estão correlacionados entre si e são ordenados de modo que os primeiros mantenham a maioria das variações presentes em todas as variáveis originais (JOLLIFFE, 1972; 1973).

Araújo (2009) afirma que o objetivo da ACP é tomar p variáveis (X_1, X_2, \dots, X_p) e encontrar combinações destas para produzir índices (Z_1, Z_2, \dots, Z_p) que sejam não correlacionados na ordem de sua importância, e que descreva a variação nos dados. Ou seja, sua finalidade básica é a análise dos dados usados visando sua redução, eliminação de sobreposições e a escolha das formas mais representativas de dados a partir de combinações lineares das variáveis originais.

A vantagem adicional da ACP está relacionada a avaliação da importância de cada característica estudada sobre a variação total, possibilitando o descarte das variáveis menos discriminantes, por estarem correlacionadas a outras variáveis, pela sua invariância ou por serem combinação linear de outras características (CRUZ e REGAZZI, 1997).

Neste contexto, alguns estudos tem mostrado a utilização da análise de componentes principais na área da ciências agrárias, utilizando essa ferramenta como mecanismo de auxílio na seleção de animais para aptidão em certas características, contribuindo na interpretação das relações existentes entre as variáveis e, conseqüentemente, na tomada de decisões (BAKER *et al.*, 1988; ROSO e FRIES, 1995).

Dentre estes estudos, agora relacionados a seleção de animais para corte, pode-se destacar o executado por Barbosa *et al.* (2005b) que objetivaram a dimensionalidade de um conjunto de variáveis referentes a carcaça de suínos e verificaram que dos 33 componentes principais, 17 (51,5%) apresentaram variância inferior a 0,7 (autovalor inferior a 0,7). Recomendou, assim, as seguintes variáveis como mais relevantes: idade ao abate; peso da meia-carcaça direita; comprimento de carcaça pelo Método Brasileiro de Classificação de Carcaça, maior espessura de toucinho na região da copa, na linha dorso-lombar, profundidade de lombo e pesos do baço, do coração, da meia-carcaça direita resfriada, do pernil, da copa, da paleta, do carré, das costelas, da papada, do filezinho e do rim.

Vale ressaltar também, o executado por Andrade (2013), que ao analisar características da carcaça de cordeiros, afirmou que para explicar as covariações das características morfométricas da carcaça e sensoriais, a realização da análise de componentes principais auxiliou na interpretação das relações existentes entre as variáveis. Com relação as características morfológicas da carcaça, este autor verificou

que o componente principal 1 obteve 74% de variação, o qual foi constituído pelas seguintes variáveis: largura do tórax, largura da garupa, profundidade do tórax, perímetro da garupa, perímetro da perna, comprimento da carcaça, comprimento interno da carcaça, comprimento da perna. Com relação ao componente principal 2, este explicou 13% da variância e foi constituído pela variável de conformação.

Ainda no âmbito da seleção de animais para corte Fernandes Júnior *et al.* (2013), buscaram avaliar por meio da ACP as características de carcaça e qualidade da carne de vários grupamentos genéticos de ovinos, e assim, através da análise exploratória dos dados, foi possível reduzir a dimensionalidade do conjunto de informações e identificar diferenças entre grupos considerando o processo formado pelas combinações lineares das características mais associadas entre si (fatores).

Constatou então que os cinco primeiros fatores extraídos por componentes principais explicaram 80 % da variância total dos dados. O primeiro fator representou as características morfométricas da carcaça, os pesos ao abate e da carcaça, os pesos dos cortes comerciais e a área de olho de lombo.

O segundo fator correspondeu ao pH medido 24 horas *post mortem*, luminosidade e intensidade de amarelo. O terceiro fator correspondeu as características espessura de gordura de cobertura e a intensidade de vermelho. O quarto fator representou as características índice de quebra por resfriamento e perda de peso por cocção, e o quinto fator foi composto apenas pela força de cisalhamento.

4 MATERIAL E MÉTODOS

a) Aquisição dos dados

Informações provenientes de um ensaio experimental, que ocorreu nas dependências do Instituto Federal de Educação, Ciência e Tecnologia do Pará (IFPA), campus Castanhal, Pará, Brasil, compuseram o banco de dados primário do estudo.

Foram utilizadas 72 carcaças de cordeiros castrados (menos de 12 meses de idade) confinados, sendo 36 Santa Inês e 36 mestiços (Dorper x Santa Inês), pertencentes à rebanhos comerciais localizados no estado do Pará, inicialmente pesados, identificados, vermifugados e instalados em baias individuais de madeira com área de 1,2 m², com o piso de concreto forrado com cama de maravalha e contendo comedouros e bebedouros individuais.

Ocorreram quatro abates aos 0, 28, 56 e 84 dias de confinamento após o período de adaptação sendo 18 animais abatidos por vez, com peso vivo médio de 35,4 kg, após 18h jejum de sólidos. A insensibilização foi feita por eletronarcose com corrente de 0,5 durante quatro segundos, seguida houve sangria, através da secção das carótidas e jugulares, ocorrendo em seguida a esfola, desarticulação das patas e do crânio e evisceração segundo os padrões convencionais determinados pela legislação vigente do Regulamento da Inspeção Industrial e Sanitária de Produtos de Origem Animal – RIISPOA (BRASIL, 2000). Em seguida as carcaças foram identificadas e pesadas para a determinação do peso da carcaça quente.

As carcaças foram resfriadas em câmara fria temperatura aproximada de 4° C permanecendo por 24 horas. Após esse período, obteve-se a carcaça fria, que foi pesada (PCF) em balança digital, identificadas, pesadas e transportadas para a câmara fria, com temperatura aproximada de 4° C, permanecendo por 24 horas. Após esse período, obteve-se o peso de carcaça fria, (PCF, em kg).

As carcaças foram classificadas de acordo com o peso e, posteriormente, tipificadas por conformação segundo o Sistema SEUROP de classificação de carcaças ovinas por um avaliador treinado através de fotografias da carcaça fria. Foram mensuradas com a utilização de paquímetro e fita métrica ambos com precisão de 0,01 cm. Tais medidas morfométricas, que serão os atributos considerados para classificação

dos escores de conformação com o KNN, foram agrupadas em dois painéis: frontal e lateral, descritas na Tabela 3 e ilustradas nas Figuras 5 A, B, C e D.

Tabela 3 – Descrição das medidas morfométricas mensuradas na carcaça de ovinos de acordo com os painéis Frontal e Lateral.

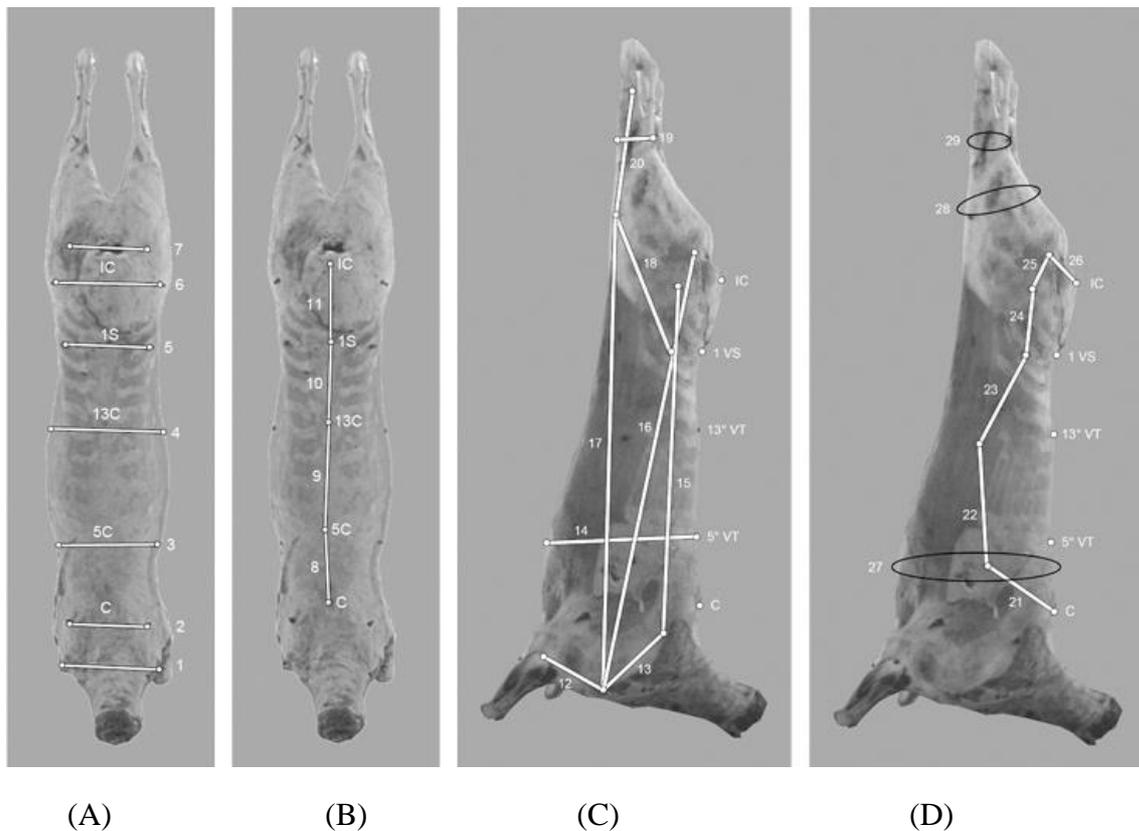
Medida	Descrição	Abreviação
Painel Frontal – Figura 5 A		
Distância entre articulações escápulo-umeral (1)	Distância entre as articulações escápulas-umerais (CAMARGO &CHIEFFI,1971).	F_DAEU
Distância entre espinhas da escápula (2)*	Distância entre o ponto mais alto das escápulas.	F_DEsE
Largura Dorsal (3)	Distância entre as faces laterais das articulações escapulo-umerais (YÁÑEZ <i>et al.</i> 2004).	F_LD
Largura lombar (4)	Distância entre os extremos da 13ª costela direita e esquerda (YÁÑEZ <i>et al.</i> 2004).	F_LL
Largura da garupa (5)	Distância entre os trocânteres maiores dos fêmures (YÁÑEZ <i>et al.</i> 2004).	F_LG
Largura entre Íleos (6)	Distância entre as pontas dos íleos.	F_LII
entre os ísquios (7)	Distância entre as pontas dos ísquios.	F_Lis
Painel Frontal – Figura 5 B		
Distância entre a cernelha à 5ª vértebra torácica (8)*	Distância entre a cernelha e a 5ª vértebra torácica.	F_DC-5T
Distância entre a 5ª vértebra torácica à 13ª vértebra torácica (9)*	Distância entre a 5ª vértebra torácica à 13ª vértebra torácica.	F_D5T-13T
Distância entre a 13ª vértebra torácica e a 1ª vértebra sacral (10)*	Distância entre a 13ª vértebra torácica e a 1ª vértebra sacral.	F_D13T-1S
Distância entre 1ª vértebra sacral e a inserção da cauda (11)*	Distância entre 1ª vértebra sacral e a inserção da cauda.	F_D1S-C
Painel Lateral – Figura 5 C		
Comprimento do braço (12)	Distância entre distância entre a articulação escápulo-umeral até a tuberosidade superior externa do rádio.	L_CB
Comprimento da escápula (13)	Distância entre o ponto mais alto da escápula à articulação escapulo-umeral (parte mais baixa) (CAMARGO &CHIEFFI,1971).	L_CE
Profundidade do Torác (14)*	Distância máxima entre o esterno e o dorso da carcaça (OLIVEIRA,2001)	L_PT
Distância entre a espinha da escápula e o ílio (15)*	Distância do ponto mais alto da escápula até o ponto mais alto do ílio.	L_DESE-II
Distância entre a articulação escápulo-umeral e o ísquio (16)*	Distância entre a articulação escápulo-umeral e o ísquio.	L_DEU-Is
Distância entre articulação escápulo-umeral e a articulação fêmur-tibial (17)*	Distância entre articulação escápulo-umeral e a articulação fêmur-tibial.	L_DAEU-FT
Comprimento da coxa (18)	Comprimento tomado da articulação do fêmur no coxal à articulação fêmur-tibial (CEZAR E SOUSA,2007).	L_CCX
Largura da perna (19)	Distância entre os bordos internos e externos da parte superior da perna em sua parte mais larga (DANTAS,2006)	L_LP
Comprimento da perna (20)	Comprimento tomado da articulação fêmur-tibial e o bordo lateral da articulação tarso-metatarsiana (YÁÑEZ <i>et al.</i> 2004).	L_CP

Painel Lateral – Figura 5 D

Distância entre a cernelha e a 5ª costela (21)*	Distância entre a cernelha e o extremo da 5ª costela.	L_DC-5Co
Distância da 5ª costela e a 13ª costela (22)*	Distância entre os extremos da 5ª e a 13ª costela	L_D5-13Co
Distância entre a 13ª costela e a tuberosidade coxal (23)*	Distância entre o extremo da 13ª costela e a tuberosidade do coxal.	L_D13C-TCx
Distância entre a tuberosidade coxal e o íleo (24)*	Distância entre a tuberosidade coxal e a ponta do íleo.	L_DTCx-IL
Distância entre o íleo e o ísquio (25)*	Distância entre as pontas do íleo e do ísquio	L_DII-Is
Distância entre o ísquio e a inserção da cauda (26)*	Distância tomada entre a ponta do ísquio e a inserção da cauda.	L_DIs-C
Circunferência do tórax (27)	Circunferência tomada do esterno a cernelha, passando-se a fita métrica por trás da paleta (SANTANA (2001)).	L_CT
Circunferência da perna (28)	Circunferência tomando como base a parte média da perna, abaixo da articulação femur-tibial	L_CP
Circunferência da coxa (29)	Circunferência tomada na base mais larga do pernil	L_CCX

* Medidas propostas pelo Programa de Seleção Genética (PROSEG) - Universidade Federal Rural da Amazônia (UFRA/ Campus Belém-PA, Brasil)

Figura 5 – Medidas Morfométricas tomadas nos painéis Frontal e Lateral.



As carcaças foram divididas longitudinalmente em duas meias carcaças com auxílio de uma serra elétrica para avaliação quantitativa do músculo *Longissimus dorsi*,

mensurado entre a 12^a e 13^a costelas, na altura do músculo (CEZAR E SOUSA, 2007). Para a obtenção da área-olho-de-lombo na carcaça (AOLCc). A mesma foi calculada por meio do *software ImageJ* 32-bit 1.6.0_24 (RASBAND, 2015), ilustrada na Figura 6 e 7.

Figura 6 - Interface do ImageJ com imagem digitalizada da área do olho-de-lombo em papel vegetal mostrando a configuração da escala, feita por meio de régua graduada em centímetros utilizada como referência para definir a distância conhecida (*Known distance*) por meio da marcação de uma reta (em amarelo).

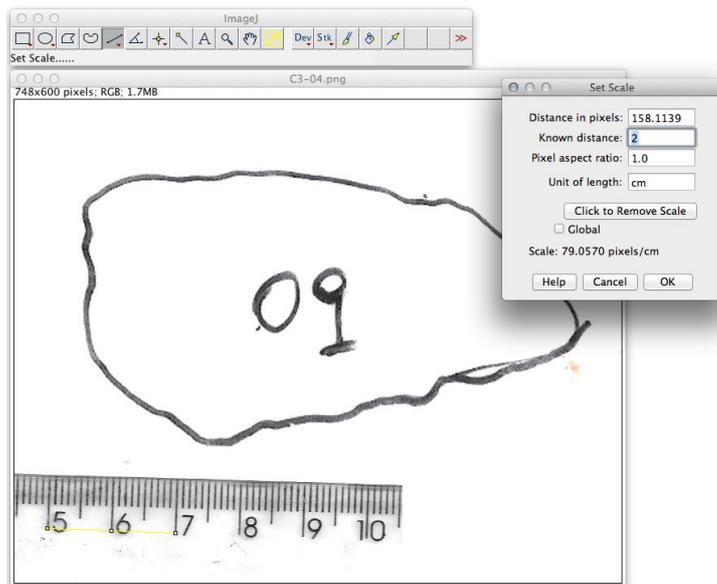
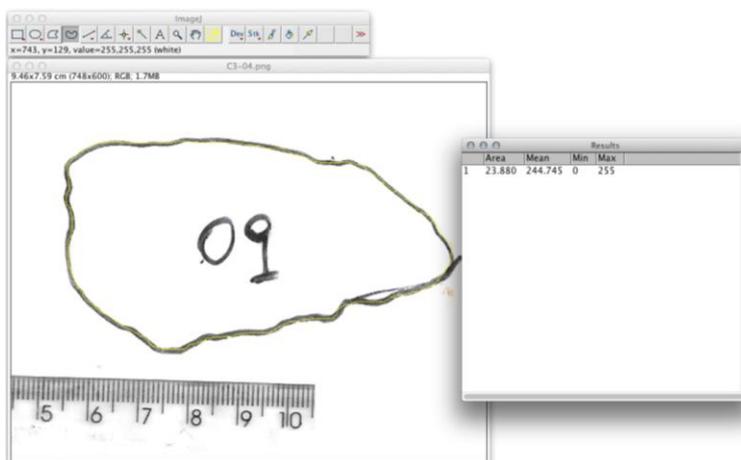


Figura 7 – Resultado da mensuração da área do olho-de-lombo, contornada em amarelo, onde o resultado é visualizado na janela *Results* na coluna *Area*.



b) Pré-processamento do banco de dados

Para a aplicação da metodologia de classificação de carcaças ovinas utilizando o KNN foram utilizadas diferentes estratégias de aprendizado como objetivo de prever a

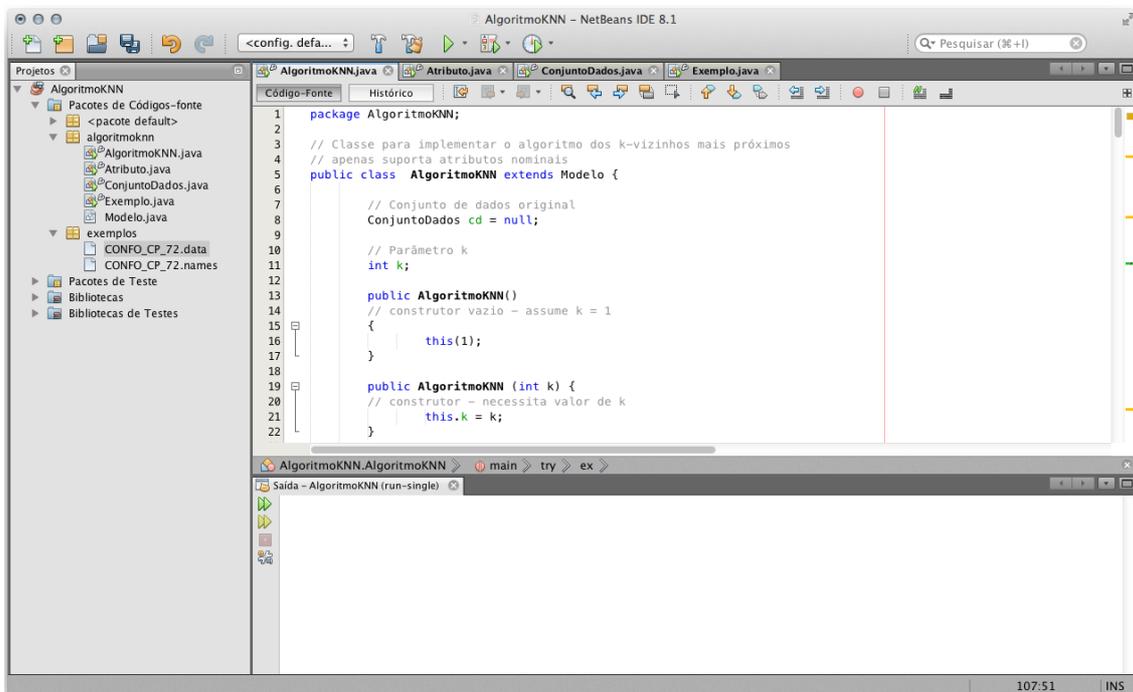
conformação das caracas ovinas. Optou-se por utilizar dois bancos de dados um de âmbito numérico e outro nominal. E conseguinte foram conduzidas várias estratégias de processamento que geraram os diferentes bancos de dados avaliados.

b.1) Pré-processamento Conjunto de dados Nominal

A ferramenta utilizada para processar as diferentes estratégias com conjunto de dados nominais foi desenvolvida em linguagem de programação Java, usando o ambiente de desenvolvimento Netbeans IDE 8.1, no sistema operacional Windows e Macintosh. As matrizes de confusão foram obtidas por script elaborado para este trabalho executado no *software R version 3.2.4* (R CORE TEAM, 2016).

Na Figura 8, esta ilustrado a Interface do NetBeans IDE 8.1, onde encontram-se os arquivos necessários à execução do algoritmo, as abas referentes às classes do programa (AlgoritmoKNN.java, Atributo.java, ConjuntoDados.java e Exemplo.java) (ROCHA et al, 2008) e a guia Saída, onde é exibido o resultado da classificação.

Figura 8 – Interface do NetBeans IDE 8.1 exibindo as guias Projetos, onde encontram-se os arquivos necessários à execução do algoritmo, as abas referentes às classes do programa e a guia Saída.



A técnica de amostragem utilizada foi a LOOCV, onde o conjunto teste é um conjunto de cardinalidade unitária, sendo considerados, alternadamente, cada um dos exemplos dados (ROCHA, 2008).

O código utilizado executado no ambiente NetBeans (Apêndice, ítem 2), classifica atributos nominais. Para tanto, foi necessário categorizar os atributos numéricos. As observações de cada variável utilizada foram dispostas em ordem crescente para cada valor x pertencente ao atributo A , a média \bar{x} e o desvio σ e estabeleceram-se nomes para os valores contidos nos seguintes intervalos: *um*, se $x < \bar{x} - \sigma$; *dois*, se $\bar{x} - \sigma \leq x < \bar{x}$; *tres*, se $\bar{x} \leq x < \bar{x} + \sigma$ e *quatro*, se $x \geq \bar{x} + \sigma$. As devidas classes de categorização das medidas estão descritas na Tabela 4.

Tabela 4 – Descrição das classes de categorização dos atributos (medidas morfométricas) provenientes da carcaça.

Atributo	Classes			
	Um	Dois	tres	quatro
Carcaça Fria (kg)	< 12,1	12,1 - 15,9	16,0 - 19,8	> 19,9
Dist. entre articulações escápulo-umeral.	< 14,5	14,5 - 16,4	16,5 - 18,4	> 18,5
Dist. entre espinhas da escápula.	< 10,6	10,6 - 12,2	12,3 - 13,9	> 14,0
Largura lombar.	< 18,1	18,1 - 20,1	20,2 - 22,2	> 22,3
Largura da garupa	< 12,5	12,5 - 13,5	13,6 - 14,6	> 14,7
Largura do ísquio.	< 11,0	11,0 - 12,5	12,6 - 14,1	> 14,2
Largura do ílio.	< 15,0	15,0 - 16,2	16,3 - 17,4	> 17,5
Distância Cernelha – 5ª vert. Torácica.	< 8,0	8,0 - 10,3	10,4 - 12,6	> 12,7
Distância 5ª vert. Torácica - 13ª vert. torácica.	< 13,3	13,3 - 15,3	15,4 - 17,4	> 17,5
Distância 13ª vert. torácica - 1ª vert. Sacral.	< 12,8	12,8 - 14,5	14,6 - 16,3	> 16,4
Distância 1ª vert. Sacral- Inserção da cauda.	< 10,5	10,5 - 12,0	12,1 - 13,7	> 13,8
Comprimento do braço (úmero).	< 11,2	11,2 - 12,5	12,6 - 14,0	> 14,1
Comprimento da escápula.	< 13,9	13,9 - 15,4	15,5 - 17,0	> 17,1
Profundidade do Toráx.	< 25,8	25,8 - 27,1	27,2 - 28,5	> 28,6
Dist. Espinha da escápula-ílio.	< 48,3	48,3 - 52,5	52,6 - 56,8	> 56,9
Dist. art. Escápula-umeral ísquio.	< 62,7	62,7 - 69,3	69,4 - 76,0	> 76,1
Dist. art. Escápula-umeral a art. Fêmur-tibial	< 75,1	75,1 - 79,0	79,1 - 83,1	> 83,2
Comprimento da coxa.	< 15,8	15,8 - 16,9	17,0 - 18,2	> 18,3
Largura da perna.	< 6,5	6,5 - 7,2	7,3 - 8,1	> 8,2
Comprimento da perna.	< 15,0	15,0 - 16,9	17,0 - 18,8	> 18,9
Distância Cernelha – 5ª Costela.	< 13,9	13,9 - 15,2	15,3 - 16,7	> 16,8
Distância 5ª Costela – 13ª Costela.	< 19,2	19,2 - 21,0	21,1 - 23,0	> 23,1
Distância 13ª Costela – tuberosidade coxal.	< 13,8	13,8 - 15,8	15,9 - 17,9	> 18,0
Distância tuberosidade coxal – íleo.	< 11,2	11,2 - 11,9	12 - 12,6	> 12,7
Distância Íleo – Ísquio.	< 5,9	5,9 - 6,8	6,9 - 7,8	> 7,9
Distância Ísquio - Inserção da cauda.	< 7,8	7,8 - 9,0	9,1 - 10,4	> 10,5
Circunferência do Tórax	< 65,5	65,5 - 70,4	70,5 - 75,4	> 75,5
Circunferência da Perna	< 16,5	16,5 - 17,7	17,8 - 19,1	> 19,2
Circunferência da Coxa	< 24,9	24,9 - 29,2	29,3 - 33,6	> 33,7
Área do olho-de-lombo (cm ²)	< 10,76	10,76 - 13,62	13,63 - 16,50	> 16,51

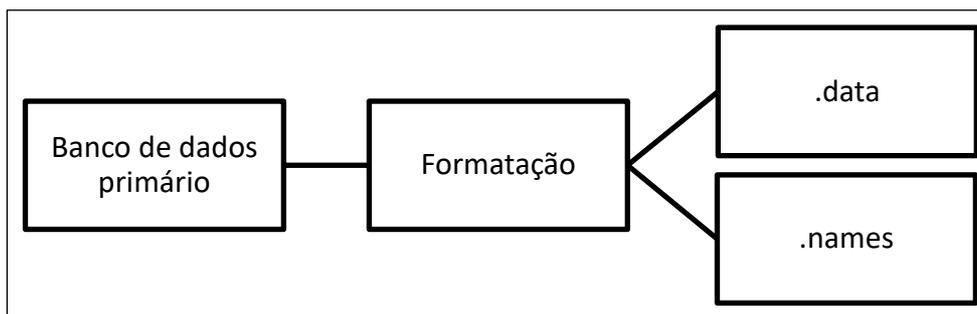
Medidas morfométricas em ‘cm’

Foram montados, então, os arquivos que alimentaram o classificador, utilizando-se duas extensões exigidas para sua execução. A extensão “.data”, contendo todos os exemplos do conjunto teste, foi formatada da seguinte maneira: cada linha correspondeu à um exemplo contendo categorias referentes à cada atributo, onde a última palavra identificava a classe (escore de conformação), escritas em letra minúscula, sem acentos, separadas por vírgulas e sem ponto final.

A extensão “.names”, contendo todos os atributos utilizados na classificação, formatados da seguinte maneira: cada linha correspondeu à um atributo, cujo nome foi escrito em minúscula, sem acentos, seguindo de espaço simples a partir do qual o conjunto de categorias possíveis identificáveis pelo classificador foi escrito sem acentos, separados por vírgula e sem ponto final.

A Figura 9, ilustra o processo de formação desses dois arquivos.

Figura 9 – Ilustração do processo de formação dos bancos de dados.



Para a formatação dos arquivos “.data” temos a exemplificação de como esse conjunto é visualizado no Netbeans ID, na Tabela 5 e Figura 10, e para os arquivos “.names” na Tabela 6 e Figura 11.

Tabela 5 – Exemplo de Estrutura das características dos bancos de dados de Conformação formam o arquivo de exemplos “.data”.

Banco de dados	Características
.data	Carcaças (exemplos)
	Características quantitativas da Carcaça (PCF)
	Medidas morfométricas
	AOLC
	Escore de Conformação

Escore de conformação: pobre, subretilínea, boa e muito boa.

Figura 10– Janela do NetBeans IDE 8.1 mostrando o arquivo CONFO_CP_72.data que contém todos os 72 exemplos (carcaças), um por linha, cujos atributos (variáveis) e a

classe (escore de conformação) são nominais com valores separados somente por vírgula.

```

53 quatro, tres, um, quatro, quatro, dois, tres, tres, quatro, tres, tres, dois, subretlinea
54 dois, dois, dois, tres, dois, quatro, dois, quatro, tres, um, dois, dois, pobre
55 um, dois, tres, dois, um, um, quatro, tres, dois, um, dois, subretlinea
56 dois, quatro, tres, dois, tres, dois, tres, tres, dois, um, um, dois, pobre
57 dois, tres, dois, tres, um, dois, dois, um, quatro, um, dois, um, subretlinea
58 um, dois, um, dois, quatro, dois, tres, tres, dois, dois, dois, dois, pobre
59 dois, dois, tres, dois, dois, dois, um, dois, dois, um, tres, dois, boa
60 quatro, tres, quatro, tres, quatro, quatro, quatro, tres, tres, tres, tres, quatro, boa
61 quatro, dois, dois, dois, tres, dois, dois, um, tres, dois, tres, tres, boa
62 quatro, quatro, tres, quatro, tres, dois, quatro, tres, tres, tres, quatro, tres, subretlinea
63 quatro, quatro, tres, quatro, dois, dois, tres, quatro, um, quatro, quatro, boa
64 tres, quatro, tres, quatro, dois, tres, dois, dois, um, quatro, tres, tres, boa
65 dois, tres, dois, dois, dois, dois, tres, tres, tres, dois, dois, pobre
66 dois, dois, quatro, dois, tres, dois, quatro, tres, tres, tres, tres, subretlinea
67 tres, dois, quatro, tres, dois, dois, dois, tres, quatro, quatro, tres, subretlinea
68 quatro, tres, tres, tres, dois, um, tres, tres, dois, quatro, quatro, quatro, muitboa
69 tres, dois, dois, dois, dois, quatro, um, um, quatro, dois, dois, boa
70 dois, dois, dois, dois, um, dois, dois, um, dois, tres, dois, subretlinea
71 tres, tres, tres, tres, quatro, tres, dois, dois, tres, tres, dois, dois, subretlinea
72 dois, tres, tres, dois, dois, dois, tres, tres, tres, dois, dois, um, subretlinea

```

Tabela 6 – Exemplo de Estrutura das características dos bancos de dados que formam o arquivo de atributos “.names”.

Atributos	Categorias
Características quantitativas da carcaça	um,dois,tres,quatro
Medidas morfométricas da carcaça	um,dois,tres,quatro
AOLC	um,dois,tres,quatro
Escore de conformação	pobre,subretlinea, boa, muito boa.

Figura 11- Janela do NetBeans IDE 8.1 mostrando o arquivo CONFO_CP_72.names que contém a descrição dos atributos (variáveis) e da classe (escore de conformação) utilizados.

```

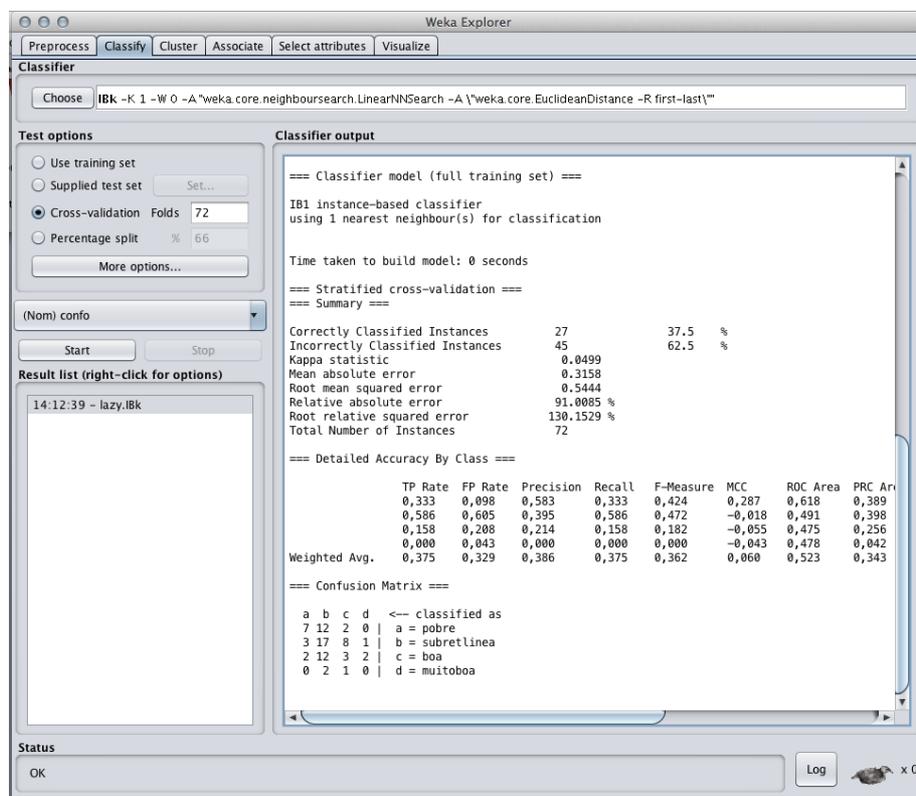
1 PCF um,dois,tres,quatro
2 M1 um,dois,tres,quatro
3 M2 um,dois,tres,quatro
4 M3 um,dois,tres,quatro
5 M4 um,dois,tres,quatro
6 M5 um,dois,tres,quatro
7 M6 um,dois,tres,quatro
8 M7 um,dois,tres,quatro
9 M8 um,dois,tres,quatro
10 M9 um,dois,tres,quatro
11 M10 um,dois,tres,quatro
12 AOL um,dois,tres,quatro
13 CONFO pobre,subretlinea,boa,muitboa

```

b.2) Pré-processamento do Conjunto de dados Numérico

Utilizou-se o *software* WEKA versão 3.8.0 para o processamento do conjunto de dados numérico. Foi configurado para executar a metodologia LOOCV, selecionando-se a opção *Cross-validation* e determinando o número de *folds* igual ao número de exemplos do respectivo conjunto, igual a 72 ou 53, da mesma maneira preconizada no código do outro *software*, utilizando os conjuntos numéricos, não categorizados, de forma a comparar o desempenho da utilização de diferentes tipos de atributos. A interface resultado dessa configuração é exemplificada na Figura 12, para k igual a 1 (saída descrita no apêndice, item 1).

Figura 12 – Interface do *software* WEKA versão 3.8.0, para o processamento do conjunto numérico.



Utilizou-se o teste de Wilcoxon para comparar a diferença entre os fatores redução ou não redução da dimensionalidade utilizando a PECC dos respectivos valores de k, à 5% de significância. O teste de normalidade de Anderson-Darling foi executado previamente para a verificação do pressuposto da normalidade, á 5% de significância. Essas análises foram executadas no *software* R *version* 3.2.4 (R CORE TEAM, 2016).

b.3) Análise de Componentes Principais

A análise de componentes principais (ACP) foi utilizada com o intuito de reduzir o número de variáveis estudadas (dimensionalidade), com menor perda possível de informação. Utilizaram-se 29 medidas morfométricas para a realização do teste normalidade dos dados e verificação de existência de *outliers* no banco de dados inicial.

Os *outliers* encontrados foram substituídos através da técnica de substituição chamada de média aparada, que consiste em excluir o maior valor e o menor valor de uma determinada variável e obter a média. Optou-se por essa alternativa para evitar a perda de informações. Ainda assim, obtiveram-se medidas não normais, que foram padronizadas pelo teste Z.

Para a realização da análise de componentes principais obteve-se a matriz de correlação entre as variáveis, para verificar quais apresentaram correlação entre 0,3 a 0,7 (MOREIRA, 2007). Tendo em vista isso, das 29 medidas utilizadas, seguindo esse critério, foram retiradas sete: Largura da perna, comprimento da escápula, profundidade externa, distância entre a articulação escapulo-umeral ao ísquio, distância entre a 5ª e a 13ª vértebra torácica, distância entre a 13ª vertebra torácica e a 1ª vértebra sacral e Distância entre o íleo e o ísquio.

Em seguida, avaliou-se a medida de adequação da amostragem Kaiser-Meyer-Olkin (KMO), de acordo com Hairet *al.*(1987), que avalia a validade da análise fatorial, onde esses valores devem ficar entre de 0,5 e 1,0. Os Intervalos de adequação amostral de KMO são classificados em: Excelente ($>0,9$) ; Meritória ($]0,8 ; 0,9]$); Intermediária ($]0,7; 0,8]$); Medíocre ($]0,6;0,7]$); Mísera ($]0,5;0,6]$) e Inaceitável ($<0,5$). No caso da análise feita neste trabalho, o valor de KMO encontrado foi de 0,891 sendo considerada meritória, recomendando assim, o uso das ACP para as seguintes análises.

A escolha do componente principal (CP) foi baseada no método sugerido por Jolliffe (1972; 1973), baseado no coeficiente de ponderação, onde o número de variáveis destacadas devem possuir autovalores superiores ou iguais a 0,7.

Após determinar o número de componentes principais (Prin) a serem utilizados, verificou-se qual medida dentre eles apresentaram maiores autovetores, sendo essas medidas as que terão maior representatividade dentro do componente. Para esta análise

utilizou-se o programa estatístico SAS (*Statistical Analysis System*) versão 8.2. As medidas morfológicas selecionadas dentro de cada componente principal.

b.4) Outras Análises Estatísticas

A descrição estatística dos atributos utilizados e coeficiente de correlação de Spearman (ρ) para cada atributo nominal com o escore de conformação foram realizadas no *software R version 3.2.4* (R CORE TEAM, 2016). Os valores de correlação foram classificados segundo Santos (2007), onde valores entre 0 e 0,19, remetem a correlação muito fraca; de 0,2 a 0,39, é considerada como fraca; de 0,4 a 0,69, moderada; 0,70 a 0,89 é classificada como forte, e; de 0,9 a 1, considerada como muito forte.

c) Classificação

Foi utilizado o Esquema de Carcaça Pesada do sistema europeu de classificação, para análises realizadas com 53 carcaças, nessa metodologia carcaças com peso de carcaça superior ou igual a 13,0 kg, são destinadas à tipificação, levando em consideração a sua conformação pelo desenvolvimento dos perfis da carcaça, nomeadamente das suas principais partes (perna, dorso e paleta).

Optou-se pela escolha de valores de k ímpares como critério de desempate, e o número de valores de k a serem testados seguiram a metodologia de Witten (2011) e Anchalia e Roy (2014).

Estabelecidas as estratégias de processamento, foi possível avaliar o desempenho do classificador em diferentes situações e, acima de tudo, o melhor desempenho do classificador, modificando o número de exemplos considerados (número de carcaças) e utilização das análises componentes principais para reduzir o número de atributos. As estratégias estão descritas a seguir:

Estratégia 1 - Utilizaram-se 72 exemplos e 31 atributos, ou seja, todas as carcaças e todas as variáveis obtidas a campo. Dessa maneira, tem-se:

CONF_NUM_72 e CONF_NOM_72

- 72 exemplos de carcaças
- 4 classes: escores de conformação da carcaça (U, R, O e P)
- 31 atributos

- 1 medida de peso da carcaça fria (PCF)
- 29 medidas lineares da carcaça fria
- 1 medida da área do olho-de-lombo (AOL)
- Atributos numéricos (NUM) ou nominais (NOM)

Estratégia 2 - Utilizaram-se 72 exemplos e 12 atributos . Sendo 10 medidas oriundas dos 10 componentes principais: largura do ísquio, largura da garupa, distância entre articulações escapulo-umeral, comprimento da perna, distância entre a cernelha e a 5º vértebra torácica, distância entre a 1º vértebra sacral e a inserção da cauda, distância entre a 5º e a 13º costelas, distância entre a tuberosidade coxal e o ílio e circunferência da perna e da coxa; peso da carcaça fria e AOLC.

CONF_NUM_72_CP e CONF_NUM_72_CP

- 72 exemplos de carcaças
- 4 classes: escores de conformação da carcaça (U, R, O e P)
- 12 atributos
 - 1 medida de peso da carcaça fria (PCF)
 - 10 medidas lineares da carcaça fria
 - 1 medida da área do olho-de-lombo (AOL)
- Atributos numéricos (NUM) ou nominais (NOM)

Estratégia 3 - Utilizaram-se 53 carcaças que apresentaram peso de carcaça fria maiores ou iguais a 13 Kg, ou seja, carcaças classificadas pelo sistema SEUROP como carcaças pesadas, e 31 atributos.

CONF_NUM_53 e CONF_NUM_53

- 53 exemplos de carcaças
- 4 classes: escores de conformação da carcaça (U, R, O e P)
- 31 atributos
 - 1 medida de peso da carcaça fria (PCF)
 - 29 medidas lineares da carcaça fria
 - 1 medida da área do olho-de-lombo (AOL)
- Atributos numéricos (NUM) ou nominais (NOM)

Estratégia 4 - Utilizaram-se 53 carcaças pesadas e os 10 componentes principais citados na Estratégia 2, peso da carcaça fria e AOLC.

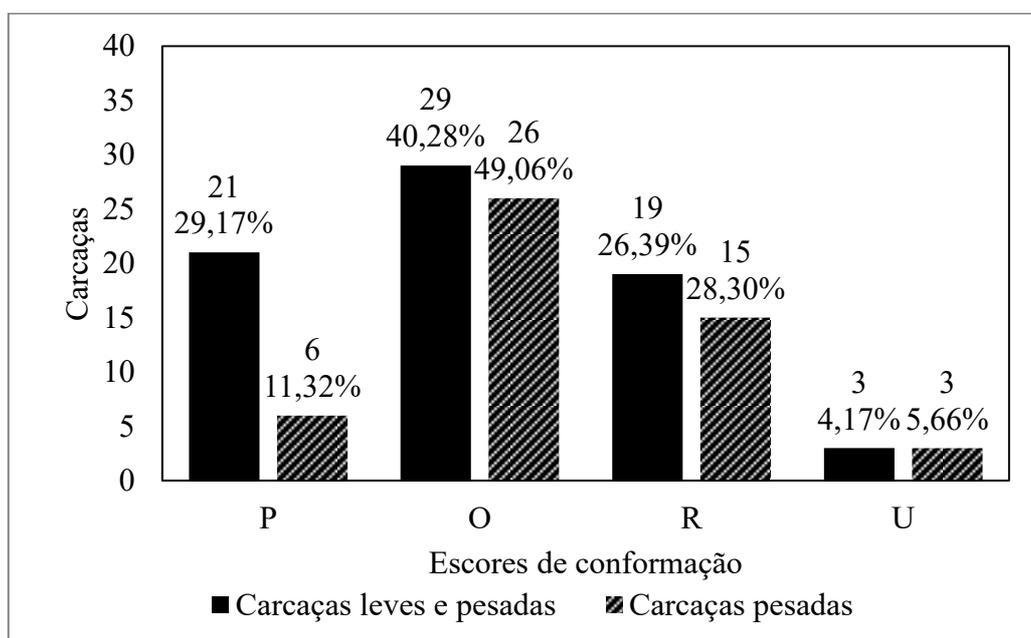
CONF_NUM_53_CP e CONF_NOM_53_CP

- 53 exemplos de carcaças
- 4 classes: escores de conformação da carcaça (U, R, O e P)
- 31 atributos
 - 1 medida de peso da carcaça fria (PCF)
 - 10 medidas lineares da carcaça fria
 - 1 medida da área do olho-de-lombo (AOL)
- Atributos numéricos (NUM) ou nominais (NOM)

5 RESULTADOS E DISCUSSÃO

Neste estudo, não foram classificadas carcaças com escore Superior (S) e Excelente (E), conforme descreve o sistema de tipificação que foi adotado. Quanto ao número e à proporção de carcaças por classe, para análises realizadas com 72 e 53 exemplos, obteve-se a distribuição e porcentagem descritas nas Figura 13. Houve uma prevalência da classe Subretlína de 40,3% e 49,1% nos dois conjuntos. Essa prevalência pode interferir na classificação já que o algoritmo KNN se baseia no número de vizinhos mais próximos ao classificar um novo exemplo.

Figura 13 – Distribuição e porcentagem relativa ao total por classe de carcaças ovinas do banco de dados de carcaças leves e pesadas (n=72) e somente com carcaças pesadas (n=53).



Na Tabela 7, vê-se a descrição estatística dos atributos utilizados e o coeficiente de correlação de cada atributo com o escore de conformação estabelecido pelo avaliador habilitado.

Os maiores valores de coeficiente de correlação (ρ) encontrados, tidos como moderados, foram dos atributos: Peso de carcaça fria (PCF) (0,65), Largura entre Íleos (F_LII) (0,60), área de olho-de-lombo (AOL) (0,56), Distância entre articulações escápulos umerais (F_DAEU) (0,53), Largura Lombar (F_LL) (0,50), Circunferência da Coxa (L_CCX) (0,49), Largura da Garupa (F_LG) (0,47) e Largura dorsal (F_LD) (0,41).

Tabela 7 – Descrição estatística dos atributos utilizados e coeficiente de correlação de cada atributo com o escore de conformação.

Atributo	Média	Desvio	Máximo	Mínimo	ρ
Painel Frontal – Figura 5 A					
Distância entre articulações escápulo-umeral (1)	16,48	1,98	20,10	9,70	0,53*
Distância entre espinhas da escápula (2)	12,28	1,68	16,90	8,70	0,35*
Largura Dorsal (3)	16,35	1,37	22,00	13,90	0,41*
Largura lombar (4)	20,21	2,10	25,80	15,50	0,50*
Largura da garupa (5)	13,60	1,09	16,00	10,80	0,47*
Largura entre Íleos (6)	16,26	1,22	18,70	12,30	0,60*
Largura entre Ísquios (7)	12,62	1,61	16,50	9,00	0,40*
Painel Frontal – Figura 5 B					
Distância entre a cernelha à 5ª vert. torácica (8)	10,36	2,36	19,30	6,30	0,17 ^{NS}
Distância entre a 5ª ver. torácica à 13ª vert. Torácica (9)	15,37	2,11	20,60	10,80	-0,15 ^{NS}
Distância entre a 13ª ver. torácica e a 1ª vert. Sacral (10)	14,62	1,80	19,60	11,50	-0,15 ^{NS}
Distância entre 1ª ver. sacral e a inserção da cauda (11)	12,13	1,67	18,00	8,20	0,15 ^{NS}
Comprimento do braço (12)	12,65	1,48	18,80	9,70	0,06 ^{NS}
Comprimento da escápula (13)	15,48	1,60	19,60	12,00	0,09 ^{NS}
Painel Lateral – Figura 5 C					
Profundidade do Torác (14)	27,18	1,42	31,00	24,20	0,28*
Distância entre a espinha da escápula e o ílio (15)	52,60	4,30	59,60	26,00	0,21 ^{NS}
Distância entre a articulação escápulo-umeral e o ísquio (16)	69,37	6,71	90,20	58,20	-0,15 ^{NS}
Distância entre articulação escápulo-umeral e a articulação fêmur-tibial (17)	79,13	4,05	88,00	60,00	0,08 ^{NS}
Comprimento da coxa (18)	17,04	1,23	20,30	14,60	0,11 ^{NS}
Largura da perna (19)	7,35	0,86	12,00	5,80	0,13 ^{NS}
Comprimento da perna (20)	16,96	1,95	20,30	9,40	0,09 ^{NS}
Painel Lateral – Figura 5D					
Distância entre a cernelha e a 5ª costela (21)	15,32	1,46	19,20	12,00	0,21 ^{NS}
Distância da 5ª costela e a 13ª costela (22)	21,13	1,97	25,10	16,50	-0,32*
Distância entre a 13ª costela e a tuberosidade coxal (23)	15,89	2,06	21,50	10,60	0,28*
Distância entre a tuberosidade coxal e o íleo (24)	11,97	0,74	14,20	10,20	-0,14 ^{NS}
Distância entre o íleo e o ísquio (25)	6,87	0,99	9,70	5,00	0,17 ^{NS}
Distância entre o ísquio e a inserção da cauda (26)	9,13	1,36	11,60	6,10	0,34*
Circunferência do tórax (27)	70,53	5,02	85,50	60,00	0,36*
Circunferência da perna (28)	17,84	1,34	20,90	14,60	0,36*
Circunferência da coxa (29)	29,32	4,40	52,20	22,00	0,49*
Peso de Carcaça Fria (Kg)	16,01	3,94	24,05	6,93	0,65*
Área de olho de lombo (cm ²)	13,63	2,88	21,29	7,41	0,56*
Escore de conformação	2,06	0,85	4,00	1,00	

^{NS} Não significativo; * P<0,05

Abreviação:

Painel Frontal 5A: 1= F_DAEU; 2= F_DEsE; 3= F_LD; 4=F_LL; 5=F_LG; 6=F_LII; 7= F_Lis

Painel Frontal 5B: 8= F_DC-5T; 9= F_D5T-13T; 10=F_D13T-1S; 11= F_D1S-C

Painel Lateral 5C: 12= L_CB; 13= L_CE; 14= L_PT; 15= L_DESE-II; 16= L_DEU-Is; 17= L_DAEU-FT; 18= L_CCX; 19= L_LP; 20= L_CP

Painel Lateral 5 D: 21= L_DC-5Co; 22= L_D5-13Co; 23= L_D13C-TCx; 24= L_DTCx-II; 25= L_DII-Is; 26= L_DIs-C; 27= L_CT; 28=L_CCX; 29=L_CP.

Ressaltando que o avaliador divide mentalmente a carcaça em três planos musculares de vista frontal: o terço anterior (cranial), o terço médio (toráx) e o terço posterior (caudal). Os valores de correlação moderado das medidas do painel frontal desse estudo, F_LII (0,60), F_LL (0,50) e F_LG (0,47), sugere que quando um avaliador habilitado realiza a tipificação por conformação, ele toma como o terço caudal como o principal terço a ser minuciosamente examinado, dada a maior valorização comercial dos cortes (Cezar e Souza, 2007).

Considerando a medida F_LG (0,47), estudos mostram que altos valores dessa medida indica uma maior proporção de músculos do corte da perna, uma característica importante a ser buscada em ovinos destinados ao abate, por apresentar maior percentual de rendimento da porção comestível, sendo considerado um dos cortes mais nobres da carcaça (PINHEIRO E JORGE, 2010).

A medida lateral L_CCX (0,49), se faz importante pois o pernil é um dos cortes de maior valor por apresentar maior percentual de rendimento da porção comestível (SOUSA, 1993). Araújo (2014), avaliando a composição tecidual dos cortes paleta e pernil de cordeiros sem raça definida, concluiu que a quantidade de músculo no pernil pôde ser estimada com eficácia pelas medidas corporais (Escore de condição corporal, altura de anterior, altura de posterior, largura de peito, largura de garupa, perímetro torácico) , que estão relacionadas com a conformação.

A determinação AOL (0,56) é considerada um bom indicador da composição corporal, Macome et al. (2008) também afirmam que devido à sua praticidade, tem sido adotada como o parâmetro mais representativo da quantidade e qualidade da carne por está diretamente ligada ao total de músculos na carcaça, obtendo importante influência na avaliação do preço final da carne e da classificação da carcaça (Cezar & Sousa, 2007)

Com a redução de atributos por meio da ACP, obtiveram-se as seguintes medidas morfométricas selecionadas dentro de cada componente principal, descritas na Tabela 8. Essas medidas compuseram as estratégias 2 e 4, tanto para os conjuntos de dados numérico quanto para o nominal.

Tabela 8 - Medidas morfométricas com maior autovetor dentro de cada componente principal selecionado.

Componente Principal	Medida morfométrica	Autovetor
Prin 1	Distância entre articulações escapulo-umeral	0,314
Prin 2	Distância entre a 5° e a 13° costelas	0,417
Prin 3	Distância entre a cernelha e a 5° vértebra torácica	0,441
Prin 4	Distância entre a 1° vértebra sacral e a inserção da cauda	0,471
Prin 5	Distância entre a tuberosidade coxal e o ílio	0,456
Prin 6	Largura do ísquio	0,349
Prin 7	Circunferência da perna	0,481
Prin 8	Largura da garupa	0,361
Prin 9	Comprimento da perna	0,300
Prin 10	Circunferência da coxa	0,386

Esses 10 componentes principais seriam responsáveis por representar a maior parcela da variação do conjunto de 29 medidas da carcaça ovina, de forma a "resumir" muitas em algumas variáveis, o que é interessante em situações experimentais por reduzir custos em tempo e pessoal.

Nessa linha, Goyache et al. (2003), comparando técnicas de predição da conformação de carcaças bovinas por meio de classificadores e regressão, selecionou 4 de 15 medidas morfométricas para representa-las cujos erros advindos da classificação foram semelhantes, ou seja, menos atributos foram necessários para que o classificador acertasse os escores.

Na Tabela 9 têm-se os resultados de exemplos corretamente classificados (PECC) por estratégia e por tipo de atributo (numérico ou nominal) com seus respectivos valores de k. Na Tabela 11, tem-se a comparação, pelo teste de Wilcoxon, entre a utilização e não utilização de redução de atributos para os diferentes conjuntos numéricos e nominais.

Nos conjuntos numéricos com exemplos de carcaças leves e pesadas (CONF_72_NUM), a Estratégia 1 (CONF_72, sem redução de atributos) teve melhor

valor de PECC de 61,11% com utilizando 7 vizinhos mais próximos. Para a Estratégia 2 (CONF_72_CP, com redução de atributos) o melhor valor de PECC foi de 62,50%, com k igual a 5. Ainda, a diferença da redução de atributos foi estatisticamente significativa ($p < 0,05$) entre as duas estratégias.

Para as carcaças pesadas (CONF_53_NUM), na Estratégia 3 (CONF_53, sem redução de atributos) foi obtido 50,94% de PECC com k igual a 5 e a 23. Para a Estratégia 4 (CONF_53_CP, com redução de atributos), foi obtido 58,49% de PECC com 11 vizinhos mais próximos.

A partir desses dados foi possível verificar que, em geral, a utilização dos esquemas de carcaça pesada (53 exemplos) diminuiu a acurácia do classificador. Para um mesmo valor de k, em sua maioria, o PECC das estratégias com esquema de carcaça pesada é menor que as estratégia que utilizam carcaças leves e pesadas. Isso pode ser devido a diminuição, principalmente, de exemplos de carcaças de score P, o que aumenta o erro de classificação dessa classe, corroborando com Witten (2011).

Nos conjuntos nominais referentes às Estratégias 1 (CONF_72_NOM), 2 (CONF_72_NOM), Estratégia 3 (CONF_53) e Estratégia 4 (CONF_53_CP) com exemplos de carcaças leves e pesadas e somente pesadas respectivamente, observou-se que o melhor parâmetro (k) observado foi o único vizinho mais próximo, cuja acurácia do classificador foi de 100% (Tabela 9). Isso indicaria que carcaças de uma mesma classe apresentam grande semelhança quanto às características morfométricas e biométricas, visualmente identificadas pelo avaliador no momento da tipificação, padrão que foi identificado pelo algoritmo.

Porém, cabe ressaltar que a distância euclidiana é uma métrica não capaz de manipular atributos nominais (BATISTA, 2003; BATISTA e SILVA, 2009). No caso do cálculo das distâncias entre valores de dois atributos nominais pela distância euclidiana, essa seria calculada em 0, se os valores fossem iguais, e 1, se os valores fossem diferentes. Sendo assim, esse resultado não reflete a realidade. Contudo não foram encontrados métricas para atributos nominais em *softwares* usuais.

De outra forma, quando se considera apenas 1 vizinho mais próximo, a flexibilidade e a variância da classificação aumentam ao passo que se utiliza apenas uma classe para a predição, o que não é interessante por desconsiderar a maioria dos

outros exemplos contidos no treinamento. Gimenez (2003) afirma que, nesse caso, a predição é irrelevante.

Como segundo melhor PECC (acurácia do algoritmo KNN), temos para a estratégia 1 e 2, o valor de k igual 5 com PECC de 66,67% e 65,28% respectivamente. Para a estratégia 3 temos 58,49% com 9 vizinhos mais próximos. E para a estratégia 4 temos 62,26% de PECC com 7 e 11 vizinhos mais próximos.

Para os dados numéricos da Estratégia 1, os valores obtidos de PECC diferiram estatisticamente ($<0,05$) da estratégia 2, que utilizou a redução de atributos, e da estratégia 3, que utilizou somente carcaças pesadas (CONF_53). Dessa maneira, um número reduzido de atributos da carcaça, 10 ao todo, conseguiu otimizar a predição (61,11% no CONF_72 versus 62,50% no CONF_72_CP).

Na Tabela 10, temos a comparação entre os fatores esquema de carcaça (leve e pesada e pesada), quantidade de atributos (todas as variáveis e ACP) e classe de dados na acurácia ou acurácia da classificação do escore de conformação de carcaças ovinas pelo teste de Wilcoxon e normalidade dos dados pelo teste de Shapiro-Wilk

Tabela 9 – Acurácia do algoritmo KNN (PECC) utilizando a metodologia de validação *cross validation leave-one-out* (LOO) para a classificação do escore de conformação de carcaças ovinas com diferentes classes de dados, dimensionalidades e número de exemplos.

K	Dados numéricos				Dados nominais			
	Carcaças leves e pesadas (CONF_72_NUM)		Carcaças pesadas (CONF_53_NUM)		Carcaças leves e pesadas (CONF_72_NOM)		Carcaças pesadas (CONF_53_NOM)	
	Estratégia 1 Sem redução de atributos (CONF_72)	Estratégia 2 Com redução de atributos (CONF_72_CP)	Estratégia 3 Sem redução de atributos (CONF_53)	Estratégia 4 Com redução de atributos (CONF_53_CP)	Estratégia 1 Sem redução de atributos (CONF_72)	Estratégia 2 Com redução de atributos (CONF_72_CP)	Estratégia 3 Sem redução de atributos (CONF_53)	Estratégia 4 Com redução de atributos (CONF_53_CP)
1	37,50%	61,11%	41,51%	50,94%	100,00%	100,00%	100,00%	100,00%
3	54,17%	47,22%	45,28%	39,62%	61,11%	63,89%	49,05%	58,49%
5	55,56%	62,50%	50,94%	49,06%	66,67%	65,28%	56,60%	60,38%
7	61,11%	58,33%	45,28%	54,72%	63,89%	54,17%	54,71%	62,26%
9	59,72%	58,33%	41,51%	56,60%	62,50%	52,78%	58,49%	50,94%
11	54,17%	61,11%	45,28%	58,49%	65,28%	55,56%	54,71%	62,26%
13	52,78%	61,11%	49,06%	54,72%	59,72%	61,11%	56,60%	50,94%
15	52,78%	61,11%	45,28%	56,50%	58,33%	61,11%	50,94%	52,83%
17	50,00%	54,17%	47,17%	54,72%	52,78%	52,78%	49,06%	50,94%
19	45,83%	55,56%	43,40%	52,83%	50,00%	51,39%	47,17%	50,94%
21	47,22%	56,94%	43,40%	52,83%	50,00%	55,56%	47,17%	47,17%
23	48,61%	55,56%	50,94%	52,83%	50,00%	55,56%	50,94%	50,94%
25	44,44%	54,17%	45,28%	54,72%	47,22%	48,61%	50,94%	47,17%

Tabela 10 - Comparação entre os fatores esquema de carcaça, quantidade de atributos e classe de dados na acurácia ou acurácia da classificação do escore de conformação de carcaças ovinas pelo teste de Wilcoxon e normalidade dos dados pelo teste de Shapiro-Wilk.

		Numérico				Nominal			
		Carcaças leves e pesadas (CONF_72_NUM)		Carcaças pesadas (CONF_53_NUM)		Carcaças leves e pesadas (CONF_72_NOM)		Carcaças pesadas (CONF_53_NOM)	
		Estratégia 1 (CONF_72)	Estratégia 2 (CONF_72_CP)	Estratégia 3 (CONF_53)	Estratégia 4 (CONF_53_CP)	Estratégia 1 (CONF_72)	Estratégia 2 (CONF_72_CP)	Estratégia 3 (CONF_53)	Estratégia 4 (CONF_53_CP)
Numérico	Carcaças leves e pesadas (CONF_72_NUM)	Estratégia 1 (CONF_72)	P>0,05						
		Estratégia 2 (CONF_72_CP)	**	P>0,05					
		Estratégia 3 (CONF_53)	**	**	P>0,05				
	Carcaças pesadas (CONF_53_NUM)	Estratégia 4 (CONF_53_CP)	P>0,05	P>0,05	**	**			
Nominal	Carcaças leves e pesadas (CONF_72_NOM)	Estratégia 1 (CONF_72)	**	**	P>0,05	P>0,05	**		
		Estratégia 2 (CONF_72_CP)	P>0,05	P>0,05	**	P>0,05	P>0,05	**	
		Estratégia 3 (CONF_53)	P>0,05	**	**	P>0,05	**	**	**
	Carcaças pesadas (CONF_53_NOM)	Estratégia 4 (CONF_53_CP)	P>0,05	P>0,05	**	P>0,05	**	P>0,05	*

Na diagonal tem-se o P-valor para o teste de normalidade de Shapiro-Wilk a 5% de significância.

Abaixo da diagonal tem-se os P-valores obtidos no teste de Wilcoxon, à 5% de significância.

Hipóteses do teste de Wilcoxon: H_0 = os dados provêm de uma mesma população; H_1 = os dados não provêm de uma mesma população.

Ainda, a retirada de exemplos teve efeito negativo na acurácia, onde 50,94% dos exemplos foram corretamente classificados no conjunto CONF_53. É possível que algumas medidas morfométricas fossem redundantes, reforçando que a dimensionalidade influencia a acurácia do algoritmo KNN.

Devido à quantidade de matrizes de confusão obtidas ser grande, $25 \times 8 = 200$ matrizes, selecionou-se aquelas pertencentes aos melhores valores de acurácia para discussão. Na Tabela 11, tem-se a matriz resultante do processamento das estratégias do conjunto numérico (CONF_72_NUM).

Tabela 11 - Matrizes de confusão para as estratégias de classificação do escore de conformação de carcaças ovinas processadas com conjuntos numéricos (CONF_72_NUM), por k de maior acurácia na estratégia.

	Estratégia 1, k = 7				Estratégia 2, k = 5			
	P	O	R	U	P	O	R	U
P	16	4	1	0	14	6	1	0
O	3	22	4	0	3	22	4	0
R	0	13	6	0	3	7	9	0
U	0	2	1	0	0	2	1	0
	Estratégia 3, k = 5				Estratégia 4, k = 11			
	P	O	R	U	P	O	R	U
P	0	5	1	0	0	6	0	0
O	1	20	5	0	0	25	1	0
R	0	11	7	0	0	12	6	0
U	0	2	1	0	0	3	0	0

Em geral, observa-se que a classe O possui a maior *recall* ou sensibilidade devido à quantidade de exemplos classificados corretamente, calculados em 75,90%, 75,90%, 76,90% e 96,2% para as Estratégias 1, 2, 3 e 4, respectivamente.

Entretanto, a precisão para essa classe é menor, calculada em 53,7%, 59,5%, 52,6% e 54,3%, respectivamente, o que se deve ao fato de o classificador generalizar incorretamente exemplos de outras classes. contrastando com Batista (2003), o qual afirmou que, frequentemente, as classes majoritárias apresentam maior precisão.

Nas Estratégias onde se utilizou apenas carcaças pesadas, a sensibilidade e a precisão para a classe P diminuiu consideravelmente em comparação às demais, 0 ficando em 0% para as Estratégias 3 e 4 contra 76,20% de precisão e 84,2% de sensibilidade para a Estratégia 1 e 70,00% de precisão e 66,70% de sensibilidade para a Estratégia 2, o que reforça o efeito negativo do desbalanço das classes.

Em sequência, o escore U não teve exemplos classificados corretamente em nenhuma das estratégias. Ele é a classe que contém apenas 3 exemplos, que foram em sua maioria classificados como sendo da classe majoritária, o escore O.

Na Tabela 12 tem-se a matriz resultante do processamento das estratégias do conjunto numérico (CONF_72_NOM). Como o maior valor de acurácia foi 100%, a sensibilidade, a acurácia e o PECC são todos absolutamente iguais a 100%. Dessa maneira, selecionou-se o segundo melhor valor de acurácia de cada estratégia para o estudo do desempenho do classificador KNN.

Tabela 12 - Matrizes de confusão para as estratégias de classificação do escore de conformação de carcaças ovinas processadas com conjuntos nominais (CONF_72_NOM), por k de maior acurácia na estratégia.

	Estratégia 1, k = 5				Estratégia 2, k = 5			
	P	O	R	U	P	O	R	U
P	20	1	0	0	19	2	0	0
O	5	21	3	0	8	16	5	0
R	2	10	7	0	1	6	12	0
U	0	1	2	0	0	1	2	0
	Estratégia 3, k = 9				Estratégia 4, k = 7			
	P	O	R	U	P	O	R	U
P	0	6	0	0	0	5	1	0
O	1	23	2	0	0	24	2	0
R	0	10	8	2	1	8	9	0
U	0	1	2	0	0	1	2	0

Para a Estratégia 1 (CONF_72), a classe P obteve grande sensibilidade e precisão, calculados em 74,07% e 95,23%, respectivamente, valores aproximados aos calculados utilizando o conjunto numérico. Esses parâmetros foram iguais a zero nas estratégias 3 e 4, onde nenhum exemplo foi classificado corretamente.

Observou-se um comportamento em geral semelhante aos da tabela anterior, onde grande parte dos exemplos foi classificado como sendo de escore O, onde se obteve 92,30% de precisão na Estratégia 4, o maior valor comparado com as demais estratégias (72,41% para a Estratégia 1, 55,17% para a Estratégia 2 e 88,46% para a Estratégia 4).

O escore R obteve sensibilidade e precisão de 58,33% e 36,84% para a Estratégia 1, 63,17% e 63,16% para a Estratégia 2, 66,67% e 40,00% para a Estratégia 3 e 64,29% e 50,00% para a Estratégia 4, respectivamente. A precisão diminuiu

consideravelmente com a retirada das carcaças com menos de 13 kg. A classe U não teve nenhum exemplo classificado corretamente.

6 CONCLUSÃO

O algoritmo KNN, foi capaz de prever os escores de conformação a partir de medidas morfométricas. A partir dos resultados da acurácia, da sensibilidade das classes e da precisão da classificação, utilizando o conjunto numérico, a melhor estratégia foi a estratégia 2 (CONF_72_CP), com redução de atributos, conseguiu otimizar a predição de classificação em relação as demais estratégias. De forma geral, o PECC obtido pelo grupo das carcaças pesadas foi menor que o grupo das carcaças leves e pesadas.

7 REFERÊNCIAS BIBLIOGRÁFICAS

_____. Portaria nº 307 de 26 dez. 1990. Aprova o "Sistema Nacional de Tipificação de Carcaças Ovinas". D.O.U., Brasília, 27 dez. 1990.

ALENCAR NÄÄS, I. Uso de técnicas de precisão na produção animal. **Revista Brasileira de Zootecnia**, v.40, p.358-364, 2011.

ALMEIDA, F. R. A., Oliveira, E. R., NASCIMENTO, I. R. **Semi confinamento de cordeiros mestiços da raça Santa Inês em pastagem de pangola (*Digitaria decumbens*), suplementados com diferentes níveis de ração concentrada.** Trabalho de Conclusão de Curso (Graduação em Engenharia Agrônômica) - Universidade Federal de Sergipe. 2002.

ANCHALIA, P. P., & ROY, K. The *k-Nearest Neighbor* Algorithm Using Map Reduce Paradigm. In Intelligent Systems, Modelling and Simulation (ISMS). **5th International Conference on (pp. 513-518) IEEE.** 2014.

ARAUJO, H. S. **Predição da composição tecidual da paleta e do pernil de cordeiros a partir de medidas in vivo e da carcaça.** Dissertação (Mestrado em Zootecnia), Faculdade de Ciências Agrárias (FCA) – Universidade Federal da Grande Dourados. 35 p. 2014.

BATISTA, G. E. A. P. A. **Pré-processamento de dados em aprendizado de máquina supervisionado.** Tese (doutorado em Ciências de Computação e Matemática Computacional) Universidade de São Paulo, 2003.

BATISTA, G. E. A. P. A.; S, D. F. How *k-Nearest Neighbor* Parameters Affect its Performance. In: 38º JAIIO - Simposio Argentino de Inteligência Artificial, 2009. Disponível em: conteudo.icmc.usp.br/pessoas/gbatista/publications.html. Acesso em: 20/01/2017.

BERGEN, R.D.; McKINNON, J.J.; CHRISTENSEN, D.A.; KOHLE, N. Prediction of lean yield in yearling bulls using real-time ultra-sound. **Canadian Journal of Animal Science**, v.76, p.305-311, 1996.

BEZERRA, S. B. L., DE CARVALHO, F. F. R., VÉRAS, R. M. L., BATISTA, Â. M. V., CABRAL, A. M. D., DE OLIVEIRA LUCAS, L. C., NETO, J. D. P. Medidas Morfométricas das Carcaças de Ovinos Submetidos a Dietas com Diferentes Níveis de Palma Forrageira. **Revista Científica de Produção Animal**, 14(2), 231-234. 2014.

BRASIL. Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa Nº3 de 17 de janeiro de 200. **Regulamento técnico de métodos de insensibilização para abate humanitário de animais de açougue.** Diário Oficial da União, Brasília, 24 jan. 2000.

BRIDI, A. M. **Normas de Avaliação, Classificação e Tipificação de Carnes e Carcaças**. Univ. Est. Londrina (2002). Disponível em <<http://docplayer.com.br/5175740-Normas-de-avaliacao-classificacao-e-tipificacao-de-carnes-e-carcacas.html>> Acessado em: 6.09.2016

CAMARGO, M.X.; CHIEFFI, A. **Ezoognósia**. São Paulo: CPA/ Instituto de Zootecnia, 1971. 320p.

CARDOSO, M. T. M., LANDIM, A. V., LOUVANDINI, H., & MCMANUS, C. (2013). Performance and carcass quality in three genetic groups of sheep in Brazil. **Revista Brasileira de Zootecnia**, 42(10), 734-742.

CARTAXO, F. Q.; SOUSA, W. H. Correlações entre as características obtidas *in vivo* por ultrassom e as obtidas na carcaça de cordeiros terminados em confinamento. **Revista Brasileira de Zootecnia**. 1490-1495. 2008.

CARVALHO, P. A.; PÉREZ, J. R.O. **Cortes comerciais em carcaças ovinas**. Departamento de Zootecnia Ovina/Universidade Federal de Lavras, 2001.

CEZAR, M. F., & SOUSA, W. D. **Carcaças ovinas e caprinas: obtenção, avaliação e classificação**. Uberaba: Editora Agropecuária Tropical. 2007.

CÉZAR, M.F; SOUSA,W. H. Proposta de avaliação e classificação de carcaças de ovinos deslanados e caprinos. **Tecnologia & Ciência Agropecuária**, João Pessoa, v.4, n.4, p.41-51, dez. 2010.

CHANDRARATNE, A, D.; KULASIRI, B.; SAMARASINGHE, S. Classification of lamb carcass using machine vision: Comparison of statistical and neural network analyses. **Journal of Food Engineering** 82, 26–34. 2007.

CLEMENTE, C,A,A. **Avaliação do desenvolvimento biométrico e rendimentos de cortes de cabritos das raças Saanen e Alpina**. Dissertação (Mestrado em Produção Animal)- Universidade Federal de Minas Gerais Escola de Veterinária, Belo Horizonte. 22 f. 2012.

CORDÃO, M. A.; CÉZAR, M. F.; SILVA, L. S.; BANDEIRA, P. A. V.;MORAES, F. F. A. Acabamento de carcaça de ovinos e caprinos-revisão bibliográfica. **Agropecuária Científica no Semiárido**, v. 8, n. 2, p. 16-23, 2012.

CORNELIO, T. C. **Características das carcaças de cordeiros alimentados com dietas contendo teores crescentes de glicerina bruta**. Tese (Doutorado)- Universidade Federal da Grande Dourados. 2014.

CUNHA, E, A.; LIMA, J. A.; SANTOS, L. E.; BUENO, M. S. Ovinocultura. In: SIMPÓSIO IZ. FEINCO 2007 DE OVINOCULTURA, 2007, São Paulo. **Anais...** São Paulo: Instituto de Zootecnia de São Paulo, 2007. p. 38-57.

DÍEZ, J., ALBERTÍ, P., RIPOLL, G., LAHOZ, F., FERNÁNDEZ, I., OLLETA, J. L., GOYACHE, F. Using machine learning procedures to ascertain the influence of beef carcass profiles on carcass conformation scores. *Meat science*, 73(1), 109-115.2006.

DUDA, R. O. **Pattern Recognition for HCI**. Department of Electrical Engineering.1996-1997.

ESTEVES, G.I.F. **Características e qualidade de carcaça de ovelhas de diferentes idades**. 2011. 55 f. Dissertação (Mestrado em Ciências Animais)-Universidade de Brasília, Brasília.

FIX, E. HODGESM J. L. DiscriminatoryAnalysis. NonparametricDiscrimination: ConsistencyProperties. *International Statistical Review*, v. 57, n. 3, 1989.

GAMA, J; CARVALHO, A. P de L.; FACELI, K; LORENA, A. C.; OLIVEIRA, M. **Extração de conhecimento de dados: data mining**. 2. ed. Lisboa: EdiçõesSílabo, 2015. 428 p.

GIMENEZ, C. M. **Identificação de bovinos através de reconhecimento de padrões do espelho nasal utilizando redes neurais artificiais**. Tese (Doutorado)- Universidade de São Paulo). 2001.

GONZAGA NETO, S.; CÉZAR, M. F.; MEDEIROS, A. N.; ARAÚJO FILHO, J. T.; PEREIRA, V. O.; COSTA, R. G. Enfoques na avaliação de carcaça ovina. Congresso Brasileiro de Zootecnia. *Anais...do ZOOTEC – Campo Grande-MS*, 2005.

GOYACHE, F., BAHAMONDE, A., ALONSO, J., LÓPEZ, S., DEL COZ, J. J., QUEVEDO, J. R., ...& DÍEZ, J. The usefulness of artificial intelligence techniques to assess subjective quality of products in the food industry.*Trends in Food Science & Technology*, 12(10), 370-381. 2001.

GUIMARÃES, F. F., SANTOS, G. V., NETO, A. A. R., BIAGIOTTI, D., SILVA SANTOS, N. P.; SARMENTO, J. L. R. Avaliação de características de carcaça por ultrassonografia em ovinos de diferentes grupos genéticos no estado do Piauí. X Simpósio Brasileiro de Melhoramento Animal. *Anais...Uberaba, MG*. 2013.

HAIR, JR. "Joseph F.; ANDERSON, Rolph E.; TATHAM, Ronald L.; BLACK, William C."**Multivariate Data Analysis: with readings**.Fourth Edition. New Jersey: Prentice-Hall, Inc (1995).

HECHENBICHLER, K., AND SCHLIEP, K.P. **Weighted k-nearestneighbor techniques and ordinal classification**. Discussion paper 399, SFB 386, Ludwig-Maximilians University, Munich. 2004

HUANG, J.; LING, C. Using auc and accuracy in evaluating learning algorithms. *Knowledgeand Data Engineering*, IEEE Transactionson, 17. v.3, p.299–310, 2005.

IMANDOUST, S. B.; BOLANDRAFTAR, M. Application of k-nearest neighbor (knn) approach for predicting economic events: theoretical background. *Int. Journal of Engineering Research and Applications*. v. 3, n. 5, 2013.

JESUS, T. I. **Considerações sobre carcaças ovinas e utilização da ultra-sonografia como ferramenta na avaliação de características de carcaças de ovinos da raça Santa Inês mantidos em confinamento para terminação**. 61p. Monografia (Graduação em Medicina Veterinária), Universidade Metodista de São Paulo, São Bernardo do Campo, 2007.

JOLLIFFE, I.T. 1972. **Discarding variables in a principal component analysis. I. Artificial data.** *Appl. Stat.*, 21(2): 160-173, 1972.

JOLLIFFE, I.T. 1973. **Discarding variables in a principal component analysis. II. Real data . . .** *APP. Stat.* 21(1): 21-31, 1973.

LAMBE, N. R., NAVAJAS, E. A., SCHOFIELD, C. P., FISHER, A. V., SIMM, G., ROEHE, R., & BÜNGER, L. The use of various live animal measurements to predict carcass and meat quality in two divergent lamb breeds. *Meat Science*, 80(4), 1138-1149. 2008.

LIPPMANN, R. **Review of neural networks for speech recognition**, *Neural Computation*, p.1-38. Keyword.: Recognition, Neural Networks. 1989.

LORENA, Ana Carolina; CARVALHO, A.C.P.L.F. Uma introdução às support vector machines. **Revista de Informática Teórica e Aplicada**, v. 14, n. 2, p. 43-67, 2007.

LOUVANDINI, H.; MCMANUS, C.; DALLAGO, B. S.; MACHADO, B. D. O.; ANTUNES, D. A. Evaluation of carcass traits, non-carcass components and 12th rib analysis of hair sheep supplemented with phosphorus. **Revista Brasileira de Zootecnia**, v. 35, n. 2, p. 550-554, 2006.

MACOME, F.M.; OLIVEIRA, R.L.; BAGALDO, A.R. MACONE, F.M.; JESUS, I.B.; GARCEZ NETO, A.F.; BARBOSA, L.P.; SILVA, M.C.A.; SOUZA, E.R.; ARAÚJO, G.G.L. Área de olho de lombo de cordeiros submetidos a dietas com torta de dendê (palmiste). In: V Congresso Nordestino de Produção Animal, 2008, Aracaju-SE. **Anais...** SNPA: Aracaju-SE. 2008. CD-ROM.

MCMANUS, C.; PAIM, T.P; LOUVANDINI, H. Avaliação Ultrassonográfica da Qualidade de Carcaça de Ovinos Santa Inês. **Ciência Animal Brasileira**, v.14, n.1, p.8-16, jan./mar, 2013.

MENEZES, J. J. L. de; GONÇALVES, H.C.; RIBEIRO, M. S.; RODRIGUES, L.; CAÑIZARES, G. I. L.; MEDEIROS, B. B. L.; GIASSETTI, A. P. Desempenho e medidas bimétricas de caprinos de diferentes grupos raciais. **Revista Brasileira de Zootecnia**. v.36, n.3, p.635-642, 2007.

MOREIRA, A.C. **Comparação da Análise de Componentes Principais e da CATPCA na Avaliação da Satisfação do Passageiro de uma Transportadora Aérea.** Departamento de Economia, Gestão e Engenharia Industrial. Universidade de Aveiro. (2007) 165-178 165.

MURTY, M. N.; DEVI, V. S. **Pattern recognition: An algorithmic approach.** Springer Science & Business Media, 2011.

OJIMA, A. L. R. O; BEZERRA, L. M. C; OLIVEIRA, A. L. R. Caprinos e ovinos em São Paulo atraem argentinos. **Instituto de Economia Agrícola.** São Paulo, 2006.

OSÓRIO, J. C. da S.; OSÓRIO, M. T. M.; Como realizar uma avaliação completa na carcaça de caprinos e ovinos. In: REUNIÃO TÉCNICA CIENTÍFICA EM OVINO-CAPRINOCULTURA, 1., 2004, Itapetinga. **Anais...** Itapetinga: Universidade Estadual do Sudoeste da Bahia, 2004. Seção Palestras. 1 CD-ROM.

OSÓRIO, J.C.S. **Estudio de localidad de canales comercializadas en el tipo ternasco según la procedencia: bases para la mejora de dicha calidad en Brasil.** Zaragoza: Universidad de Zaragoza, 1992. 335p. Tese (Doutorado em Veterinária) - Universidad de Zaragoza, 1992.

OSÓRIO, J.C.S.; OSÓRIO, M.T.M.; OLIVEIRA, N.R.M.; SIEWERDT, L. Qualidade, Morfologia e Avaliação de Carcaças. Universidade Federal de Pelotas, Ed. Universitária. 194p. 2002.

PALSSON, H. Meat qualities in sheep with special reference to Scottish breed and sample joint as indice of quality and composition. **Journal of Agricultural Science** , v.29, p.544-626, 1939.

PANDORFI, H., ALMEIDA, G. L. P., & GUISELINI, C. Zootecnia de precisão: princípios básicos e atualidades na suinocultura. **Revista Brasileira de Saúde e Produção Animal**, 13 v.2. 2012.

PAREDES, R.; VIDAL, E. **Learning Weighted Metrics to Minimize Nearest-Neighbor Classification Error.** IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.28, no.7, pp. 1100-1110, 2006.

PÉREZ, J. R. O., CARVALHO, P. A., PAULA, O. J. **Aspectos relacionados com a produção de carne ovina.** UNESP– Grupo de Nutrição de Ruminantes. 2008.

PINHEIRO, R. S. B., & JORGE, A. M. Medidas biométricas obtidas in vivo e na carcaça de ovelhas de descarte em diferentes estágios fisiológicos. **Revista Brasileira de Zootecnia**, 440-445. 2010.

PINHEIRO, R.S.B. **Aspectos quantitativos da carcaça e qualitativos da carne de ovinos de diferentes categorias.** 105f. Dissertação (Mestrado em Zootecnia) - Faculdade de Ciências Agrárias e Veterinárias/Universidade Estadual Paulista, Jaboticabal. 2006.

PINHEIRO, R.S.B.; JORGE, A.M. Medidas biométricas obtidas in vivo e na carcaça de ovelhas de descarte em diferentes estágios fisiológicos. **Revista Brasileira de Zootecnia**, v.39, n.2, p.440-445, 2010.

PIRES, C.C.; CARVALHO, S.; GRANDI, A. D. Características quantitativas e composição tecidual da carcaça de cordeiros terminados em confinamento. **Ciência Rural**, v.29, n.3, p.539-543, 1999.

PRICE, M. A. Development of carcass grading and classification systems. In: JONES, S. D. M. (Ed.) Quality and grading of carcasses of meat animals. **Boca Raton: CRC Press**, p.173-199. 1995.

REGULAMENTO (CE) n. 22/2008 de la Comisión Europea, 11 de Enero de 2008, Disposiciones de Aplicación del Modelo Comunitario Europeu de Clasificación de Canales de Ovino. Madrid, 2008.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações**. Barueri: Editora Manole, 2005.

RICARDO, H. A.; TAVARES, S. A. ; SURGE, C. A. ; ANDRADE, E. N. ; ROCA, R. O. . Classificação de carcaças ovinas pelo modelo EUROP: peso vivo e escore de condição corporal de animais com menos de doze meses de idade. In: 38º Congresso Brasileiro de Medicina Veterinária, 2011, Florianópolis. **Anais... 38º CONGRESSO BRASILEIRO DE MEDICINA VETERINÁRIA**, 2011. v. 1.

RICARDO, H.A. **Desenvolvimento de padrões para avaliação e classificação comercial de carcaças ovinas**. Tese (Doutorado) – Universidade Estadual Paulista, Faculdade de Medicina Veterinária e Zootecnia, Botucatu, 2011. 79p.

ROCHA, M., CORTEZ, P. & Neves, J. **Análise Inteligente de Dados – Algoritmos e Implementação em Java**. 1. ed. Brasil: FCA – Editora de Informática, 2008.

ROSO, V.M.; FRIES, L.A. Componentes principais em bovinos da raça Polled Hereford à desmama e sobreano. **Rev. Soc. Bras. Zootec.**, v.24, p.728-735, 1995.

SAINZ , R. D.; ARAUJO, F. R. C. **Tipificação de carcaças de bovinos e suínos**. Animal Science Dept.University of California, Davis, CA, USA. 2001.

SAINZ, R. D. Avaliação de carcaças e cortes comerciais de carne caprina e ovina. In: SIMPÓSIO INTERNACIONAL SOBRE CAPRINOS E OVINOS DE CORTE, 1., 2000, João Pessoa, PB. **Anais...** João Pessoa: [s.n.], 2000. p. 237-250.

SANTANA, A.F., MARTINS FILHO, R. Fatores que influenciam no desenvolvimento ponderal de ovinos jovens deslançados. **Arquivos da Escola de Medicina Veterinária da UFBA**, Salvador, v.18, n.1, p.41-60, 1995/1996.

SANTOS, C. **Estatística descritiva: manual de Auto-aprendizagem**. Lisboa: Sílabo, p.312, 2007.

SILVA, I. J. O. **Contribuições à zootecnia de precisão na produção industrial de aves e suínos no brasil.** Escola Superior de Agricultura “Luis de Queiroz”. Universidade de São Paulo. 2007.

SILVA, L. F.; PIRES, C. C.. Avaliações quantitativas e predição das proporções de osso, músculo e gordura da carcaça de ovinos. **Revista Brasileira de Zootecnia**, v.29, n.4, p.1253-1260, 2000.

SILVA, S. L. **Estimativa de características de carcaça e ponto ideal de abate por ultrassonografia, em bovinos submetidos a diferentes níveis energéticos na ração.** Diss. Universidade de São Paulo (USP). Faculdade de Zootecnia e Engenharia de Alimentos, 2002.

SOUSA, O. C. R. **Rendimento de carcaça, composição regional e física da paleta e quarto em cordeiros Romney Marsh abatidos aos 90 e 180 dias de idade.** 102f. Dissertação (Mestrado em Zootecnia) - Universidade Federal de Pelotas, Pelotas, 1993.

SOUSA, W., OJEDA, M., MORAIS, O., & ROCHA, J. Programa de melhoramento genético para ovinos deslanados e caprinos de corte do Brasil: uma proposta. **Tecnol. & Ciên. Agropec.**, João Pessoa, v.2, n.3, p.41-46, set. 2008.

SUGISAWA, L.; MATTOS, W.R.S.; OLIVEIRA, H.N. *et al.* Correlações simples entre as medidas de ultra-som e a composição da carcaça de bovinos jovens. **Revista Brasileira de Zootecnia**, v.35, n.1, p.169-176, 2006

SYED, M. E. **Attributeweighting in K-nearestneighbor classification.** Dissertação (mestrado). Escola de Ciências da Informação - School of Information Sciences. Universidade de Tampere, 50 p.2014

THEODORIDIS, S.; KOUTROUMBAS, K. **Pattern Recognition**, Second Edition. Orlando, FL, USA: Academic Press, 2003.

TODESCO, José L. **Reconhecimento de Padrões usando uma rede neuronal artificial com uma função de base radial: uma aplicação na classificação de cromossomos humanos.** Tese (Doutorado) Universidade Federal de Santa Catarina, Florianópolis. 1995.

UNIÃO EUROPÉIA. Regulamento (CE) N.º 22/2008 da Comissão, de 11 de janeiro de 2008, que estabelece as regras de execução da grelha comunitária de classificação das carcaças de ovinos. Jornal Oficial, n.º L 9 de 12/01/2008, p. 6-11.

VIANA J. G. A. **Panorama Geral da Ovinocultura no Mundo e no Brasil.** **Revista Ovinos**, Ano 4, N.º 12, Porto Alegre, Março de 2008. Disponível em <<http://www.almanaquedocampo.com.br/imagens/files/panorama%20geral%20ovinocultura%20brasil.pdf>> Acesso em: 19. julho.2016.

WANG, B.; TANG, H.; GUO, C.; XIU, Z. Entropy optimization of scale-free networks robustness to random failures. **Physica: Statistical Mechanics and its Applications**, v.363, n.2, páginas 591-596, 2006.

WANG, J.; NESKOVIC, P.; COOPER, L. N. Improving nearest neighbor rule with a simple adaptive distance measur. **Pattern Recognition Letters**, n. 28, 2007.

WANGENHEIM, C. G.; WANGENHEIM, A. **Raciocínio Baseado em Casos**. Barueri, SP: Manole, 2004. 293 p. p. 7. ISBN 85-204-1459-1.

WEBB, G. I., J. BOUGHTON, AND Z. WANG . Not So Naive Bayes: Aggregating One-Dependence Estimators. **Machine Learning** 58(1). Netherlands: Springer, pages 5-24. 2005

WITTEN, I. H.; EIBE, F.; HALL, M. A. **Data mining: practical machine learning tools and techniques**. 3º ed. Elsevier: 2011. 665 p.

WONG, Wai Kit *et al.* Secure kNN computation on encrypted databases. In: Proceedings of the 2009 ACM SIGMOD International Conference on Management of data. ACM, 2009. p. 139-152.

YAMAMOTO, S. M. **Desempenho e características de carcaça de cordeiros terminados em confinamento com dietas contendo silagem de peixe**. 2006. 95 f. Tese (Doutorado em Zootecnia) – Universidade Estadual Paulista, 2006.

YÁÑEZ, E. A.; RESENDE, K. T.; FERREIRA, A. C. D.; MEDEIROS, A. N.; SILVA SOBRINHO, A. G.; PEREIRA FILHO, J. M.; TEIXEIRA, I. A. M.A.; ARTONI, S. M. B. Utilização de Medidas Biométricas para Predizer Características da Carcaça de Cabritos Saanen. **Revista Brasileira de Zootecnia**, v.33, n.6, p.1564-1572,2004.

ZUNDT, M.; MACEDO, F. A. F.; Nunes Martins, E.; MEXIA, A. A.; NIETO, L. M.; YAMAMOTO, S.M.; MACEDO, R.M.G. Características de carcaça de cordeiros terminados em confinamento, com dietas contendo diferentes níveis protéicos. **Ciênc. rural**, v. 33, n. 3, p. 565-571, 2003.

APÊNDICE

1. Saída gerada pelo *software* WEKA, para k igual a 7 e metodologia LOOCV.

=== Run information ===

```
Scheme:          weka.classifiers.lazy.IBk -K 1 -W 0 -A
"weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R
first-last\""
Relation:        CONFO_NUM
Instances:       72
Attributes:      32
                 PCF
                 LIS
                 LIL
                 LGA
                 LLO
                 LTO
                 DEE
                 DAEU
                 CPE
                 LPE
                 CCO
                 CBR
                 CES
                 PEX
                 DEE_IL
                 DAEU_IS
                 DAEU_AFT
                 DC_5VT
                 D5VT_13VT
                 D13VT_1VS
                 D1VS_IC
                 DC_5C
                 D5C_13C
                 D13C_TC
                 DTC_IL
                 DIL_IS
                 DIS_IC
                 PT
                 PP
                 PCO
                 AOL
                 confo
Test mode:       72-fold cross-validation
```

=== Classifier model (full training set) ===

```
IB1 instance-based classifier
using 1 nearest neighbour(s) for classification
```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	27	37.5	%
Incorrectly Classified Instances	45	62.5	%
Kappa statistic	0.0499		
Mean absolute error	0.3158		
Root mean squared error	0.5444		

```

Relative absolute error          91.0085 %
Root relative squared error     130.1529 %
Total Number of Instances       72

```

=== Detailed Accuracy By Class ===

Area	PRC Area	Class	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC
0,389		pobre	0,333	0,098	0,583	0,333	0,424	0,287	0,618
0,398		subretlinea	0,586	0,605	0,395	0,586	0,472	-0,018	0,491
0,256		boa	0,158	0,208	0,214	0,158	0,182	-0,055	0,475
0,042		muitoboa	0,000	0,043	0,000	0,000	0,000	-0,043	0,478
0,343		Weighted Avg.	0,375	0,329	0,386	0,375	0,362	0,060	0,523

=== Confusion Matrix ===

```

a b c d <-- classified as
7 12 2 0 | a = pobre
3 17 8 1 | b = subretlinea
2 12 3 2 | c = boa
0 2 1 0 | d = muitoboa

```

2. Código em Java utilizado no ambiente Netbeans IDE.

```

package AlgoritmoKNN;

// Classe para implementar o algoritmo dos k-vizinhos mais próximos
// apenas suporta atributos nominais
public class AlgoritmoKNN extends Modelo {

    // Conjunto de dados original
    ConjuntoDados cd = null;

    // Parâmetro k
    int k;

    public AlgoritmoKNN()
    // construtor vazio - assume k = 1
    {
        this(1);
    }

    public AlgoritmoKNN (int k) {
    // construtor - necessita valor de k
        this.k = k;
    }

    public void construirModelo (ConjuntoDados cd)
    // apenas guarda a referência ao conjunto de dados
    {
        this.cd = cd;
    }

    public int similaridade(Exemplo ex1, Exemplo ex2) {
    // calcula similaridade entre dois exemplos
        int res = 0;
        for(int i=0; i< ex1.numValores()-1; i++)

```

```

        if(ex1.valorAtributo(i) == ex2.valorAtributo(i))
            res++;
    return res;
}

public int[] construirListaVizinhos (Exemplo ex) {
// constroi uma lista com os k vizinhos mais proximos de um exemplo
    int [] vizinhos = new int[this.k]; // lista de vizinhos
    int [] sims = new int[this.k]; // lista de similaridades
    int minsim = cd.numAtributos(); // valor mais alto possivel
    int minindex=0;
    for(int i=0; i< this.k; i++) { // coloca os primeiros k exemplos
        vizinhos[i] = i;
        sims[i] = similaridade(ex, cd.exemplo(i));
        if(sims[i] < minsim) {
            minsim = sims[i];
            minindex = i;
        }
    }

    for(int i=k; i< cd.numExemplos(); i++) {
        int s = similaridade(ex, cd.exemplo(i));
        if(s > minsim) { // substitui o menor dos que estavam por um
melhor
            vizinhos[minindex] = i;
            sims[minindex] = s;
            // recalcula o minimo
            minsim = sims[0];
            minindex = 0;
            for(int j=1; j<k; j++)
                if(sims[j] < minsim) {
                    minsim = sims[j];
                    minindex = j;
                }
            }
        }
    return vizinhos;
}

public double prever(ConjuntoDados cd, Exemplo ex) {
// classifica um novo exemplo
    this.cd = cd;
    int [] vizinhos = construirListaVizinhos(ex);
    // calcula classe mais comum nos vizinhos
    double cmc = -1;
    int maxex = 0;
    for(int c=0; c < cd.saida().numValores(); c++) {
        int nec = 0;
        for(int i=0; i< k; i++)
            if( cd.exemplo(vizinhos[i]).valorSaida()==c )
                nec++;
        if (nec > maxex) {
            maxex = nec;
            cmc = c;
        }
    }
    return cmc;
}

// Main: um exemplo !!
public static void main(String [] args)
{
    try{

```

```

        /* código main original */
        // O valor entre parênteses corresponde aos k-vizinhos mais
próximos avaliados.
        // O exemplo deve conter o mesmo número de variáveis!
        //ConjuntoDados cd = new
ConjuntoDados("/Users/iMacLacib06/AlgoritmoKNN/src/exemplos/CONFO_72");
        //AlgoritmoKNN kn= new AlgoritmoKNN (1);
        //kn.construirModelo(cd);
        //Exemplo ex = new
Exemplo("dois,tres,tres,tres,dois,quatro,tres,dois,dois,dois,dois,um,tres,dois,d
ois,dois,dois,dois,quatro,tres,tres,tres,tres,quatro,tres,tres,tres,dois,dois,do
is,um,subretlinea", cd.atributos());
        //System.out.println("Classe para o exemplo:" +
kn.classificar(cd,ex) );

        //código main alternativo, comentado:
        ConjuntoDados cd = new
ConjuntoDados("/Users/iMacLacib06/AlgoritmoKNN/src/exemplos/CONFO_72");
        AlgoritmoKNN kn= new AlgoritmoKNN(4);
        kn.construirModelo(cd);

        System.out.println("PECC:" + kn.pecc(cd, cd.exemplos() ) );
        System.out.println("Matriz de confusao:" + kn.pecc(cd,
cd.exemplos() ) );
        int [][] matriz = kn.matriz_confusao(cd, cd.exemplos() );
        for (int i=0; i< matriz.length; i++) {
            for (int j=0; j< matriz.length; j++)
                System.out.print( matriz[i][j] + " " );
            System.out.println(""); }

    }
    catch(Exception E){
        E.printStackTrace();
    }
}
}
}

```