

UNIVERSIDADE ESTADUAL DO MARANHÃO
CENTRO DE CIÊNCIAS TECNOLÓGICAS
PÓS-GRADUAÇÃO EM ENGENHARIA DA COMPUTAÇÃO E SISTEMAS

TIAGO SANTOS FERREIRA

**CLASSIFICAÇÃO DIAGNÓSTICA DE CASOS DE LEISHMANIOSE CANINA
UTILIZANDO APRENDIZADO DE MÁQUINA**

São Luís

2021

TIAGO SANTOS FERREIRA

**CLASSIFICAÇÃO DIAGNÓSTICA DE CASOS DE LEISHMANIOSE CANINA
UTILIZANDO APRENDIZADO DE MÁQUINA**

Dissertação apresentada como requisito para a
obtenção do grau de mestre em Engenharia de
Computação.

Área: Telemática e Automação

Orientador: Prof. Dr. Ewaldo Eder Carvalho
Santana

São Luís

2021

Ferreira, Tiago Santos.

Classificação diagnóstica de casos de leishmaniose canina utilizando aprendizado de máquina / Tiago Santos Ferreira. – São Luís, 2022.

44

Dissertação (Mestrado) – Programa de Pós Graduação em Engenharia de Computação e Sistemas, Universidade Estadual do Maranhão, 2022.

Orientador: Prof. Dr. Ewaldo Éder Carvalho Santana.

1.Aprendizado de máquina. 2.Classificação. 3.Regressão logística.
4.Leishmaniose visceral canina. I.Título.

CDU: 636.7:[616.993.161:004.85]

TIAGO SANTOS FERREIRA

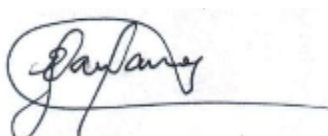
**CLASSIFICAÇÃO DIAGNÓSTICA DE CASOS DE LEISHMANIOSE CANINA
UTILIZANDO APRENDIZADO DE MÁQUINA**

Dissertação apresentada como requisito para a
obtenção do grau de mestre em Engenharia de
Computação.

Área: Telemática e Automação

Aprovado em: 16/06/2022

BANCA EXAMINADORA



Dr. Ewaldo Eder Carvalho Santana (Orientador)
Universidade Estadual do Maranhão



Dr. Mauro Sérgio Silva Pinto (Examinador Interno)
Universidade Estadual do Maranhão



Dr. Paulo Fernandes da Silva Júnior (Examinador Externo)
Universidade Estadual do Maranhão



Dr. Nilviane Pires Silva Sousa (Examinador Externo)
Faculdade Laboro

AGRADECIMENTOS

Agradeço aos meus pais Cássia Raquel Santos e Paulo Henrique de Jesus Ferreira, pelo apoio, amor e dedicação; por batalharem, para me dar as melhores condições possíveis. Aos meus Irmãos Marcos Henrique Santos Ferreira e Paulo Vitor Santos Ferreira por sempre estarem ao meu lado.

Agradeço ao meu orientador Dr. Ewaldo Santana, pelos ensinamentos, pela oportunidade de trabalhar no Laboratório de Aquisição e Processamento de Sinais, e pelo tempo dedicado a me orientar. Ao Dr. Paulo Fernandes, por me auxiliar na pesquisa deste trabalho e revisão do mesmo, além disso, pela inspiradora dedicação e paixão pela ciência.

Agradeço aos meus amigos, Andre Leite, Laís Barros, Jessica Lauane, Nathalia Lima, Leonardo Tavares, Ana Carolina, Caroline Tavares, Raiane Gonçalves, pelo amor, risadas, aventuras e apoio; por serem tal qual uma família pra mim.

Agradeço a todos da Serhum Consultoria que me apoiaram nesta jornada, Aparecida Bessa, Camila Bessa, Sarah Penha, Sâmia Brito, Kelle Menezes, Alexsandro Feitosa, Erick Pereira, Alana Bastos, Marcos Costa, Jefferson Sousa, Amparo Sousa, Luiza Pinto.... E tantos outros que não caberiam em uma só página, obrigado por estarem comigo.

“Tudo é natureza.”

Ailton Krenak

RESUMO

Resumo: Propostas de técnicas que reduzam custos financeiros no diagnóstico e tratamento de doenças animais são bem-vindas. Este trabalho utiliza algumas técnicas de aprendizado de máquina para classificar se há ou não casos de leishmaniose visceral canina por meio de exames físicos, dessa forma, sendo uma ferramenta para auxiliar na triagem desses animais em áreas endêmicas, se tornando útil para auxiliar profissionais da área da saúde na tomada de decisão. Para validação do método, foram escolhidos quatro modelos de aprendizado de máquina: K-nearest neighbor, Naïve Bayes, Máquina de Vetor de Suporte e o modelo de Regressão Logística. Os testes foram realizados em trezentos e quarenta cães, utilizando-se dezoito características do animal e o teste sorológico ELISA (*Enzyme-Linked Immunosorbent Assay*) como validação. A regressão logística obteve as melhores métricas: acurácia de 75%, sensibilidade de 84%, especificidade de 67%, razão de verossimilhança positiva de 2,53 e razão de verossimilhança negativa de 0,23, mostrando uma relação positiva na avaliação entre a verdadeira positivos e rejeitando os casos de falsos negativos.

Palavras-chave: aprendizado de máquina; classificação; regressão logística; leishmaniose visceral canina.

ABSTRACT

Abstract: *Techniques that reduce financial costs in the diagnosis and treatment of animal diseases are always welcome. This work uses different machine learning techniques to classify whether or not there are cases of canine visceral leishmaniasis through physical examinations, thus being a tool to assist in the screening of these animals in endemic areas, becoming useful to assist professionals in decision making. For validation of the method, four machine learning models were chosen: K-nearest neighbor, Naïve Bayes, support vector machine and logistic regression models. The tests were performed on three hundred and forty dogs, using eighteen characteristics of the animal and the ELISA (enzyme-linked immunosorbent assay) serological test as validation. Logistic regression achieved the best metrics: Accuracy of 75%, sensitivity of 84%, specificity of 67%, a positive likelihood ratio of 2.53 and a negative likelihood ratio of 0.23, showing a positive relationship in the evaluation between the true positives and rejecting the cases of false negatives.*

Keywords: *machine learning; classification; logistic regression; canine visceral leishmaniasis*

LISTA DE FIGURAS

Figura 1 - Número de casos novos de LV por UF em 2016.	19
Figura 2 - Modelo SVM linear. Duas classes (vermelho versus azul) foram classificadas.	24
Figura 3 - Gráfico da curva logística onde $\alpha=0$ e $\beta=1$.	27
Figura 4 - Região do Maranhão, Brasil, onde os dados foram coletados.	29
Figura 5 - Curva ROC para aplicação do modelo LR no conjunto de testes.	34

LISTA DE TABELAS

Tabela 1 - Estatísticas descritivas e associações univariáveis de características incluídas na previsão de aprendizado de máquina da leishmaniose visceral canina (casos: n = 177; não casos: n = 163).	30
Tabela 2 - Indicações da matriz de confusão.	33
Tabela 3 - Métricas de teste dos modelos. Pode-se ver que a Regressão Logística obteve as melhores métricas.	34
Tabela 4 - Matriz de confusão.	36
Tabela 5 - Frequência das variáveis para as amostras de falsos negativos.	36
Tabela 6 - Frequência das variáveis para as amostras falso-positivas.	37
Tabela 7 - Frequência das variáveis para as amostras de verdadeiros negativos.	38
Tabela 8 - Frequência das variáveis para as amostras verdadeiras positivas.	39

SUMÁRIO

1 INTRODUÇÃO.....	10
1.1 Objetivos.....	11
1.1.1 Objetivo geral.....	11
1.1.2 Objetivo específico.....	11
2 FUNDAMENTAÇÃO TEÓRICA E ESTADO DA ARTE.....	12
2.1 Estado da Arte.....	12
2.2 Leishmaniose Visceral Canina.....	13
2.3 Diagnóstico Da LVC.....	17
2.4 Aprendizado de Máquina.....	19
2.4.1 Naive Bayes.....	21
2.4.2 K Nearest Neighbor.....	22
2.4.3 Support Vector Machine.....	22
2.4.4 Regressão Logística.....	24
3 MATERIAIS E MÉTODOS.....	27
3.1 Seleção Das Variáveis E Modelo De Aprendizagem.....	29
3.2 Teste Diagnóstico.....	31
4 RESULTADOS E DISCUSSÕES.....	32
5 CONSIDERAÇÕES FINAIS.....	37
REFERÊNCIAS.....	39

1 INTRODUÇÃO

Técnicas de Aprendizado de Máquina (AM) empregam o princípio de inferência denominado indução, no qual obtêm-se conclusões a partir de um conjunto particular de exemplos. O aprendizado indutivo pode ser dividido em dois tipos principais: supervisionado e não-supervisionado. No aprendizado supervisionado tem-se a figura de um “professor” externo, o qual apresenta o conhecimento do ambiente por conjuntos de exemplos na forma: entrada e saída desejada. O algoritmo de AM extrai a representação do conhecimento a partir desses exemplos. O objetivo é que a representação gerada seja capaz de produzir saídas corretas para novas entradas não apresentadas previamente.

A regressão logística é uma das ferramentas analíticas mais utilizadas nas ciências sociais e naturais. A regressão logística é o algoritmo de aprendizado de máquina supervisionado de linha de base para classificação, que possui um relacionamento próximo com as redes neurais artificiais (RNA). A regressão logística pode ser usada para classificar uma observação em uma dentre duas, ou mais classes.

As leishmanioses constituem um complexo de enfermidades diferenciadas, que podem acometer as mucosas, a pele ou as vísceras, de acordo com o tipo de espécie envolvida e da resposta imunológica do hospedeiro (WHO, 2010).

A Leishmaniose Visceral (LV) é uma doença com amplo espectro clínico é considerada grave, especialmente nos casos humanos não tratados, podendo apresentar altos índices de letalidade. A LV está presente em 88 países, sendo 22 nas Américas. No continente Americano, estima-se que o Brasil seja responsável por 90% dos casos de LV na América Latina (NASCIMENTO et al, 2005).

Sobre o Maranhão, Furtado et al. (2015) relatam expansão de casos de leishmaniose humana no período de 2000 a 2009. Sendo que do ano de 1999 a 2005, o estado liderou em número de casos confirmados da doença no Brasil. Essa liderança é confirmada em 2016, segundo os dados do SINAN/SVS/MS, pois o Maranhão apresentou 655 casos confirmados.

O Maranhão tem o segundo pior Índice de Desenvolvimento Humano (IDHM) do país. O índice é calculado com base nos dados de 2010, leva em conta a expectativa de vida, a escolaridade e a renda da população (WHO, 2013). Pode-se afirmar que, atualmente o IDHM, é um dos indicadores urbanos que melhor avalia a qualidade de vida nas cidades.

De acordo com Catão (2012), há uma inter-relação estável entre o patógeno, vetores e pessoas (infectadas e susceptíveis) com o espaço geográfico. No caso da leishmaniose, os mapeamentos podem auxiliar na compreensão da dinâmica da transmissão e no comportamento dos vetores. Os agravos podem ser restritos a uma localidade ou atingir áreas mais extensas.

Assim, o conhecimento dos padrões espaciais na ocorrência da doença torna-se importante para a vigilância dos casos (SIQUEIRA, 2009).

A microrregião da Baixada Maranhense pertence à mesorregião do Norte Maranhense e está localizada a Oeste do Estado do Maranhão ($1^{\circ}59'-4^{\circ}00'S$ e $44^{\circ}21'-45^{\circ}33'W$). Está dividida em 21 municípios, possui uma área de 17.579, 366 km² e uma população estimada em 563.877 habitantes. Abrange os Municípios de Anajatuba, Arari, Bela Vista do Maranhão, Cajari, Conceição do Lago-Açu, Igarapé do Meio, Matinha, Monção, Olinda Nova do Maranhão, Palmeirândia, Pedro do Rosário, Penalva, Perimirim, Pinheiro, Presidente Sarney, Santa Helena, São Bento, São João Batista, São Vicente Ferrer, Viana e Vitória do Mearim. Sua população é predominantemente rural, com exceção nos municípios Arari, Pinheiro, Santa Helena, São Bento e Viana que apresentam a população urbana mais expressiva, segundo o censo do Instituto Brasileiro de Geografia e Estatística (IBGE, 2016).

Estudar a dinâmica espaço-temporal das vulnerabilidades ambientais da leishmaniose visceral humana e a prevalência da doença canina na Baixada Maranhense é de grande importância, uma vez que possibilita compreender como os processos de contágio da doença vêm se estabelecendo ao longo do tempo e do espaço. Neste contexto, esta pesquisa pretende auxiliar no avanço de políticas públicas voltadas ao controle da leishmaniose, bem como na ampliação de conhecimento científico e tecnológico para a região.

Portanto, pela falta de estudos relacionados vulnerabilidades dos casos de leishmaniose humana e leishmaniose visceral canina na Baixada Maranhense, pela contribuição sobre a relação entre a saúde humana e a qualidade de vida em ambientes rurais, além de ressaltar a importância da doença no debate sobre a qualidade de vida da população dos municípios da baixada maranhense, faz-se necessário este estudo acerca da LV nesta região.

Neste trabalho utilizamos o algoritmo de Regressão Logística, um método clássico de aprendizado supervisionado, para inferir a presença ou não da Leishmaniose Visceral em Cães (*Canis lupus familiaris*) na região da baixada maranhense.

1.1 Objetivos

1.1.1 Objetivo geral

- Analisar e Desenvolver um algoritmo de aprendizagem de máquina capaz de diagnosticar a Leishmaniose Visceral em cães.

1.1.2 Objetivo específico

- Elucidar os conceitos empregados nos métodos de aprendizagem de máquina utilizado;

- Expor acerca da leishmaniose canina, sua presença no Brasil e no estado do Maranhão;
- Demonstrar que a inteligência artificial é capaz de identificar com certo grau de precisão os animais que apresentam a doença.

2 FUNDAMENTAÇÃO TEÓRICA E ESTADO DA ARTE

2.1 Estado da Arte

Vários estudos usaram o aprendizado de máquina para diagnosticar doenças caninas. Larios, G. et. al. (2021) desenvolveram um método para o diagnóstico da leishmaniose visceral canina baseado em espectroscopia de infravermelho com transformada de Fourier (espectroscopia FTIR) e aprendizado de máquina, no qual foram analisados soros sanguíneos caninos de vinte cães não infectados, vinte *Leishmania infantum* e oito cães infectados com *Trypanosoma evansi*. Eles usaram análise de componentes principais com algoritmos de aprendizado de máquina e arquivaram mais de 85% no diagnóstico de verdadeiros positivos. Reagan, KL et al. (2020) também aplicaram técnicas de aprendizado de máquina para auxiliar no diagnóstico de Hipoadrenocorticismo Canino (HC) usando o diagnóstico de triagem por hemograma completo e painel de química sérica. A base de dados utilizada foi de 908 cães controle com suspeita de HC e 133 cães com HC confirmado. Um algoritmo de árvore orientada foi treinado e testado para avaliar o desempenho, com sensibilidade de 96,3% e especificidade de 97,2%.

Um modelo de previsão de carga parasitária linfonodal a partir de dados clínicos em cães com leishmaniose visceral por redes neurais artificiais e aprendizado de máquina foi apresentado em Torrecilha, P. et. al. (2017). Neste estudo, 55 (cinquenta e cinco) cães de sete regiões dos estados da Bahia, Minas Gerais, São Paulo e Distrito Federal, sendo 35 cães infectados e vinte cães controles, arquivaram acurácia de 78% nas análises realizadas. Na pesquisa realizada em Schofield, I. et. al. (2021), quatro algoritmos de aprendizado de máquina foram usados para prever o diagnóstico da síndrome de Cushing, usando dados clínicos estruturados do programa VetCompass no Reino Unido. A síndrome de Cushing, que é uma doença endócrina em cães, afeta negativamente a qualidade de vida dos cães afetados. Métodos de aprendizado de máquina podem classificar os diagnósticos de síndrome de Cushing registrados, com resultado preditivo para regressão com sensibilidade de 0,71 e especificidade de 0,82. Podemos notar que nesses trabalhos todos os pesquisadores utilizaram algum tipo de exame laboratorial.

2.2 Leishmaniose Visceral Canina

A leishmaniose pertence ao grupo de doenças causadas por um protozoário parasita do gênero *Leishmania*, sendo transmitida para os humanos e outros vários mamíferos através da picada de fêmeas de um inseto hematófago – dípteros da família *Psychodidae*, subfamília *Phlebotominae*, conhecidos genericamente como flebotomíneos, fazendo papel de vetor no ciclo da doença (NEVES, 2016, WHO, 2010). Neste trabalho, é abordada a formas visceral.

A leishmaniose visceral (LV) ou calazar é uma zoonose primariamente de canídeos silvestres e domésticos, causada por parasitas do gênero *Leishmania* (REY, 2001; DIETZE & CARVALHO, 2003; MICHALICK & GENARO, 2005).

As primeiras referências à possível existência de leishmaniose visceral na América do Sul são de Carlos Chagas que, percorrendo o vale do Rio Amazonas e seus principais afluentes, entre 1911 e 1912, suspeitou da ocorrência da doença nesta região por encontrar pacientes com esplenomegalia sem causa justificada. Alguns anos depois, a doença foi identificada também na Argentina, por Mazza, que em 1926 relatou os primeiros casos autóctones da doença em humanos na província de Salta. No entanto, somente em 1934, com os estudos de Penna, é que se confirmou que a doença, até então considerada inexistente do ponto de vista epidêmico, era autóctone e de alta ocorrência nas Américas. Já no ano de 1936, enquanto um novo foco epidêmico surgia na província de Chaco, Argentina, Evandro Chagas realizou, por meio de punção esplênica, o primeiro diagnóstico in vivo da doença, classificando a *Leishmania chagasi* como uma nova espécie do gênero *Leishmania* (CHAGAS et al., 1936).

Muitos fatores podem ter contribuído para a dispersão geográfica da LV no Brasil, entre eles o movimento de cães entre áreas endêmicas e não endêmicas, e mudanças na ecologia do vetor. O vetor da LV está disseminado pelo Brasil e encontra-se adaptado para colonizar o meio ambiente modificado pelo homem. Isso significa que a introdução de cães infectados em áreas não endêmicas, onde existam potenciais vetores, podem resultar em um novo foco da doença (DANTAS-TORRES et al., 2006).

A LV apresenta distribuição mundial, sendo causada por distintas espécies de *Leishmania* no Velho e no Novo Mundo. É uma doença infecciosa grave que acomete as vísceras e que pode ser fatal se não tratada, pois o hospedeiro falha em montar uma resposta protetora eficiente contra o parasita. Os primeiros casos de leishmaniose visceral aconteceram na Índia no ano de 1885 e, somente alguns anos mais tarde, em 1903, é que o agente causador desta enfermidade foi descoberto e descrito por William Boog Leishman e Charles Donovan. William B. Leishman descreveu o parasita, mas associando-o às formas de *Trypanosoma* (LEISHMAN, 1903). Charles Donovan encontrou o parasitário no baço de uma criança hindu

com febre irregular, mas o confundiu com outro protozoário, o *Trypanosoma brucei*. Após algumas descrições equivocadas, Ronald Ross criou o gênero *Leishmania* e batizou o agente causador do calazar de *Leishmania donovani*, em homenagem a William Boog Leishman e Charles Donovan (PESSOA & MARTINS, 1988; REY, 2001).

A possível participação dos cães no ciclo epidemiológico do calazar começou a ser aventada por Nicolle e Comte em 1908, na Tunísia, a partir da detecção, nos animais, do agente etiológico do calazar (NICOLLE & COMTE, 1908). A primeira demonstração de infecção natural de hamsters por picada de flebotomíneo ocorreu em 1931, após observação de que a distribuição do vetor *Phlebotomus argentipes* era coincidente com a distribuição da doença (ADLER, & THEODOR, 1931). O primeiro caso humano autóctone brasileiro foi descrito no Paraguai proveniente do estado do Mato Grosso (MIGONE, 1913). Penna em 1934 encontrou pela primeira vez o parasita em lâminas histológicas de fígado de pacientes com febre amarela oriundos da região norte e nordeste (PENNA, 1934). No Brasil, uma das primeiras observações da infecção canina por *Leishmania* foi realizada por Evandro Chagas quando demonstrou a existência da doença no homem e no cão e a infecção do flebótomo *Lutzomyia longipalpis*. O parasita foi classificado como *Leishmania chagasi* (CHAGAS, 1936; CHAGAS et al., 1938).

Atualmente a leishmaniose visceral está presente em todos os continentes, com exceção à Antártida e a Oceania, sendo a estimativa global do número de casos humanos de LV ficando em torno de 500.000 por ano e de mortalidade é de 50.000 a 200.000 mortes anualmente, especialmente em épocas de epidemia (WHO, 2010). 90% dos casos estão restritos à Índia, Bangladesh, Nepal, Sudão e Brasil. O Brasil é responsável por 90% dos casos de leishmaniose visceral nas Américas. A doença começou a ter notificação obrigatória no Brasil em 1978 e, ultimamente, cerca de 70% dos estados da federação brasileira já registraram casos autóctones do calazar (FUNASA, 1999). A região Nordeste representa o principal foco da doença, com 90% dos casos, sendo os estados da Bahia, Piauí, Ceará e Maranhão, responsáveis por estes valores (FUNASA, 1999). Geralmente a LV está associada às áreas semi-áridas, sendo originalmente uma doença típica de áreas rurais e periurbanas (GONTIJO & MELO, 2004). O que se observa normalmente é que a leishmaniose visceral canina (LVC) antecede a doença humana. Cães de áreas endêmicas podem apresentar aspecto clínico compatível com a LVC, sendo importante o diagnóstico seguro visto a expansão ininterrupta da doença. Nenhum teste de diagnóstico da LVC apresenta 100% de sensibilidade e de especificidade. Detectar o cão soropositivo, removê-lo e sacrificá-lo são medidas ainda duvidosas enquanto estratégia de controle. O diagnóstico clínico da leishmaniose visceral canina é precário e complexo, pois os sinais clínicos da doença são variáveis e inespecíficos, comuns a outras enfermidades que

acometem o cão. A imunossupressão causada por *Leishmania* pode gerar infecções oportunistas, dificultando do mesmo modo o diagnóstico da leishmaniose visceral (SILVA, 2007).

O cão doméstico (*Canis familiaris*) desempenha grande importância na epidemiologia da LV em áreas endêmicas, pois é o único reservatório doméstico da LV. Estima-se que aproximadamente 2,5 milhões de cães na Europa estejam infectados, e que na América do Sul o número de cães infectados também esteja na casa dos milhões, com as maiores taxas de infecção em países como o Brasil e a Venezuela (BANETH & SOLANO-GALLEGU, 2012). Outros animais podem compor o cenário epidemiológico da LV em áreas endêmicas, pois, mesmo com a retirada de para todos os cães soropositivos, o ciclo de transmissão da doença ainda pode existir, embora o número de casos humanos tenha diminuído (ASHFORD et al., 1998).

O papel do cão como o mais importante reservatório da *L. chagasi* em áreas endêmicas tem sido contestado, atribuindo também ao homem a responsabilidade de ser um reservatório deste parasito no Brasil (COSTA, 1997; COSTA et al., 2000).

A infecção em cães por espécies de *Leishmania* é clinicamente semelhante à infecção humana, embora no cão, além do acometimento das vísceras, são frequentemente encontradas lesões de pele nos animais infectados e sintomáticos (KRAUSPENHAR et al., 2007).

O quadro clínico é variável e depende da resposta imune do cão e da cepa do parasita inoculado pela picada do inseto vetor (MICHALICK & GENARO, 2005). Inicialmente surge febre intermitente, perda de peso e linfadenopatia (LIMA et al., 2004). Alguns cães curam espontaneamente enquanto que outros evoluem até a morte em poucas semanas (MICHALICK & GENARO, 2005).

A importância do cão na epidemiologia da doença não reside somente no fato do mesmo apresentar altas prevalências de infecção quando comparadas à espécie humana, mas também pelo elevado número de animais assintomáticos, que pode chegar a 80% da população infectada (DANTAS-TORRES et al., 2006; BANETH et al., 2008; PALTRINIERI et al., 2010). Esses servem de fonte de infecção para o vetor e, muitas vezes, deixam de ser identificados numa população devido à ausência de sintomas, ou ainda, em função de resultados falso-negativos nos exames sorológicos (BANETH et al., 2008). Em alguns cães os sinais clínicos da doença aparecem logo após a infecção, porém, em muitos animais a infecção segue seu curso de forma assintomática. No primeiro caso, os cães desenvolvem imunidade humoral, podem ser identificados por sorologia, mas são incapazes de desenvolver uma imunidade celular efetiva.

Por outro lado, os cães que não exibem sintomas da doença podem permanecer desta forma por anos ou por toda a vida. No entanto, uma alteração em seu estado imune, em decorrência de alguma enfermidade ou do uso de medicamentos imunossupressores, pode levar ao aparecimento dos sintomas da doença (BANETH et al., 2008). O desenvolvimento de sintomas inespecíficos e, muitas vezes, tardios contribui para o subdiagnóstico da doença e transmissão do parasita no ambiente doméstico. A alta prevalência de cães assintomáticos sugere que esses animais mantêm o ciclo de transmissão da LV na mesma proporção, ou até em proporção superior aos cães sintomáticos.

Apesar de existirem discordâncias no que diz respeito à importância epidemiológica de cães assintomáticos, recentemente foi comprovado que cães assintomáticos são altamente competentes para estabelecer a infecção em flebotomíneos, demonstrando o seu papel na manutenção do ciclo epidemiológico da doença (LAURENTI et al., 2013). O uso de técnicas moleculares para o diagnóstico da LVC, particularmente em animais soronegativos, permitiu a confirmação de que quando as condições são favoráveis para a transmissão da doença (tais como altas densidades de animais), a infecção se dissemina rapidamente pela população canina. Mesmo com a disseminação da infecção em áreas endêmicas, nem todos os cães infectados desenvolvem a doença. Esses conceitos demonstram que os casos clínicos de LVC, nessas áreas, são apenas o topo de um iceberg, uma vez que a maioria da população está exposta e torna-se infectada sem apresentar evidências clínicas da doença ou anticorpos anti-*Leishmania* sp (BANETH et al., 2008).

Características individuais ou biológicas, tais como a idade e o sexo, podem ser um fator de risco para a infecção por *Leishmania* sp. em cães. A maior predisposição em machos tem sido observada (DANTAS-TORRES et al., 2006). A LV pode acometer animais de qualquer idade, entretanto, a enfermidade em cães caracteriza-se por ocorrer de forma bimodal. O primeiro pico ocorre em animais com menos de três anos de idade e o segundo mo que estão na faixa entre oito a dez anos (PALTRINIERI et al., 2010). Dantas-Torres, Brito e Brandão Filho (2006), observaram que no Estado de Pernambuco, os cães jovens, com idade inferior a um ano, apresentaram alto risco de infecção por *L. infantum chagasi*. O estado imunológico ainda imaturo do animal, associado às condições ambientais, pode ser um fator determinante. Além disso, em áreas endêmicas ocorre a reposição de cães submetidos à eutanásia, ou que vieram a óbito, por animais jovens, o que torna a população canina mais susceptível a diversas enfermidades infecciosas, entre elas a LV (ANDRADE et al., 2007).

A LVC é uma doença crônica, fatal e sistêmica, sendo os principais sinais clínicos no cão representados pela caquexia, hipergamaglobulinemia, hepatoesplenomegalia, anemia e

linfadenopatia (CIARAMELLA et al., 1997; FERRER, 1999; LIMA et al., 2004; BRITO et al., 2004; LANGONI et al., 2005; LINHARES et al., 2005; KRAUSPENHAR et al., 2007). Na pele são comuns úlceras crostosas na orelha, focinho e região periorbital, descamação furfurácea e alopecia multifocal. As preparações citológicas de pele de orelha podem demonstrar a presença de formas amastigotas de leishmanias (CIARAMELLA et al., 1997; LINHARES et al., 2005; KRAUSPENHAR et al., 2007).

Lesões oculares como querato conjuntivite, blefarite, inflamação mononuclear plasmocitário do trato uveal e edema de córnea, formação de sinéquia, lesões em corpo ciliar e íris podem estar associado a depósito de imunocomplexos nestas áreas, fato que pode ser corroborado pela presença de anticorpos específicos anti leishmania em vários tecidos intra-oculares, podendo significar lesões de origem imunopatológica (GARCIA-ALONSO et al., 1996; CIARAMELLA et al., 1997; FERRER, 1999; BRITO et al., 2004).

2.3 Diagnóstico Da LVC

Diagnosticar clinicamente a LVC é um problema para as autoridades de saúde, pois existe um amplo espectro de sinais e sintomas. A doença pode estar clinicamente inaparente por anos, dependendo da fase da doença e do estado imunológico do animal, em torno de 60% a 80% dos cães que vivem em áreas endêmicas podem ter contato com o parasito e não desenvolver sinais clínicos da doença (QUEIROZ, 2010). Entretanto, já foi demonstrado que cães infectados assintomáticos podem transmitir para flebotomíneos, tendo um papel ativo na transmissão da doença (ALVAR et al., 1994; MOLINAR et al., 1994; BRANCHES et al., 1998). Nesta situação, o cão pode se constituir fonte de infecção para os insetos vetores, pois mesmo aparentemente sadios podem conter formas do parasita, sobretudo na pele, além do baço, fígado e linfonodos superficiais. (SILVA, 2007)

O diagnóstico é realizado por exames RIFI (Reação de Imunofluorescência Indireta - *Indirect Immunofluorescence assay*), ELISA (Ensaio Imunoenzimático - *Enzyme Linked Immunosorbent Assay*), PCR (Reação em Cadeia pela Polimerase - *Polymerase Chain Reaction*), teste rápido e observação direta da forma amastigota do protozoário pela análise de esfregaço (DANTAS-TORRES et al., 2017). Com o tratamento, ocorre melhora clínica e diminuição da carga parasitária no animal, mas não o torna livre de ser um reservatório e não impede que os sinais clínicos não voltem (BANETH et al., 2002).

No Brasil, as estratégias de controle da doença estão baseadas no diagnóstico e tratamento precoce de casos humanos, no controle dos vetores por meio do uso de inseticidas e na detecção dos cães infectados por análises parasitológicas e sorológicas RIFI e ELISA, com

posterior eutanásia dos animais positivos (BRASIL, 2003); por isso um diagnóstico de baixo custo e alta confiabilidade é necessário para a triagem desses reservatórios. Com a intenção de se evitar o uso de métodos invasivos e considerando que a resposta humoral na LVC é geralmente muito intensa, com altos níveis de imunoglobulinas, o diagnóstico passou a ser focado em métodos sorológicos (ALVAR, 2004). Assim, diversos métodos sorológicos foram desenvolvidos e são bastante utilizados para o diagnóstico canino.

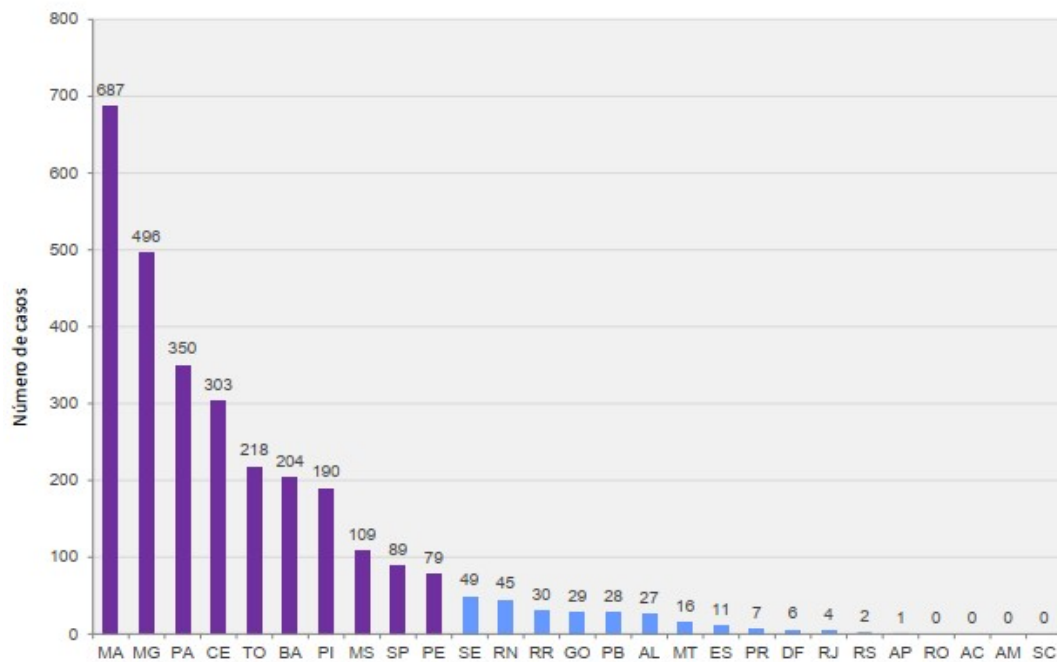
O Ensaio Imunoenzimático é considerado o padrão-ouro dos imunoensaios. (ALHAJJ, 2021). O é um método sorológico quantitativo constituindo uma ferramenta utilizada tanto para análise de suspeita clínica quanto para confirmação do diagnóstico de leishmaniose. A confirmação ocorre através da detecção de imunoglobulina G (IgG) no soro de cães suspeitos. Esse exame é escolhido por sua especificidade e sensibilidade (ALVES, 2009) (FONSECA et al., 2018). O desempenho do teste ELISA está relacionado não apenas ao tipo de antígeno utilizado, mas também ao estado clínico manifestado pelo cão (FARIA et al., 2012). Neste trabalho, os resultados do teste ELISA (positivo ou não) são utilizados como variável dependente (alvo).

A prevenção requer a conscientização das pessoas sobre a gravidade da doença para que se possam tomar medidas no controle do vetor para proteção dos animais nessas áreas, com uso de repelentes, vacinas e pulverização do ambiente (WERNECK et al., 2002).

A Organização Mundial de Saúde incluiu as leishmanioses como uma das seis doenças mais importantes do mundo. Mesmo incluídas nessa condição, as leishmanioses são consideradas doenças negligenciadas, diretamente relacionadas com a pobreza dos indivíduos infectados, às condições ruins de habitação e saneamento básico, e o baixo índice de desenvolvimento da região observada (CRUZ, 2016).

Sobre o Maranhão, Furtado et al. (2015) relatam expansão de casos de leishmaniose humana no período de 2000 a 2009. Sendo que do ano de 1999 a 2005, o estado liderou em número de casos confirmados da doença no Brasil. Essa liderança é confirmada em 2016, segundo os dados do SINAN/SVS/MS, pois o Maranhão apresentou 655 casos confirmados (DE LIMA JUNIOR, 2018) como pode observado na Figura 1.

Figura 1 - Número de casos novos de LV por UF em 2016.



Fonte: DE LIMA JUNIOR (2018).

Diante da importância dos conhecimentos de epidemiologia para o entendimento de como o processo saúde doença ocorre em um território e do fornecimento de informações para as tomadas de decisões e direcionamento das medidas de controle e prevenção, se faz necessário aprofundar o conhecimento sobre a leishmaniose visceral canina na Região da baixada maranhense.

2.4 Aprendizado de Máquina

Em vez de codificar o conhecimento em computadores, o aprendizado de máquina (ML) busca aprender automaticamente relacionamentos e padrões significativos a partir de exemplos e observações (BISHOP, 2006). Os avanços no ML permitiram o recente surgimento de sistemas inteligentes com capacidade cognitiva semelhante à humana que penetram em nossos negócios e vida pessoal e moldam as interações em rede nos mercados eletrônicos de todas as maneiras possíveis, com as empresas aumentando a tomada de decisões para produtividade, engajamento e funcionários. retenção (SHRESTHA et al. 2021), sistemas assistentes treináveis que se adaptam às preferências individuais do usuário (FISCHER et al. 2020) e agentes de negociação que abalam os mercados tradicionais de negociação financeira (JAYANTH BALAJI et al. 2018).

A capacidade de tais sistemas para resolução avançada de problemas, geralmente denominada inteligência artificial (IA), é baseada em modelos analíticos que geram previsões, regras, respostas, recomendações ou resultados semelhantes. Alimentados pela praticidade de

novas estruturas de programação, disponibilidade de dados e amplo acesso ao poder de computação necessário, os modelos analíticos vêm sendo construídos utilizando o que geralmente é chamado de Aprendizado de Máquina (*Machine Learning* – ML) (BRYNJOLFSSON & MCAFEE 2017; GOODFELLOW et al. 2016).

Em termos gerais, a IA compreende qualquer técnica que permita aos computadores imitar o comportamento humano e reproduzir ou superar a tomada de decisão humana para resolver tarefas complexas de forma independente ou com intervenção humana mínima (RUSSELL & NORVIG 2021). Como tal, está preocupado com uma variedade de problemas centrais, incluindo representação do conhecimento, raciocínio, aprendizado, planejamento, percepção e comunicação, e se refere a uma variedade de ferramentas e métodos; por exemplo, raciocínio baseado em casos, sistemas baseados em regras, algoritmos genéticos, modelos fuzzy, sistemas multiagentes (CHEN et al. 2008). As primeiras pesquisas de IA se concentraram principalmente em declarações codificadas em linguagens formais, sobre as quais um computador pode raciocinar automaticamente com base em regras de inferência lógica. Isso também é conhecido como abordagem de base de conhecimento (GOODFELLOW et al. 2016). No entanto, o paradigma enfrenta várias limitações, pois os humanos geralmente lutam para explicar todo o seu conhecimento tácito necessário para realizar tarefas complexas (BRYNJOLFSSON & MCAFEE 2017).

O aprendizado de máquina supera essas limitações. De um modo geral, ML significa que o desempenho de um programa de computador melhora com a experiência em relação a alguma classe de tarefas e medidas de desempenho (JORDAN & MITCHELL, 2015). Como tal, visa automatizar a tarefa de construção de modelos analíticos para realizar tarefas cognitivas como detecção de objetos ou tradução de linguagem natural. Isso é alcançado pela aplicação de algoritmos que aprendem de forma iterativa a partir de dados de treinamento específicos do problema, o que permite que os computadores encontrem insights ocultos e padrões complexos sem serem explicitamente programados (BISHOP, 2006). Especialmente em tarefas relacionadas a dados de alta dimensão, como classificação, regressão e *clustering*, o ML mostra boa aplicabilidade. Aprendendo com cálculos anteriores e extraíndo regularidades de bancos de dados massivos, pode ajudar a produzir decisões confiáveis e repetíveis. Por esse motivo, os algoritmos de ML têm sido aplicados com sucesso em muitas áreas, como detecção de fraudes, pontuação de crédito, análise da próxima melhor oferta, reconhecimento de voz e imagem ou processamento de linguagem natural.

O objetivo geral do Aprendizado de Máquina, é reconhecer padrões em dados, que informam o caminho invisível de problemas a serem tratados. Por exemplo, em um ambiente

altamente complexo como um carro autônomo, grandes quantidades de dados provenientes de sensores têm que ser transformados em decisões de como controlar o carro por um computador que “aprendeu” reconhecer o padrão de “perigo” (CARLEO et al, 2019).

O sucesso do ML nos últimos tempos tem sido marcado a princípio por melhorias significativas em algumas tecnologias, por exemplo, no domínio do reconhecimento de imagens. Em grande medida, esses avanços constituíram as primeiras demonstrações do impacto que os métodos de ML podem ter em tarefas especializadas. Mais recentemente, aplicativos tradicionalmente inacessíveis ao software automatizado foram habilitados com sucesso, em particular pela tecnologia de Aprendizado Profundo (*Deep Learning*). A demonstração do aprendizado por reforço técnicas de jogo, por exemplo, teve um profundo impacto na percepção de que todo o campo estava se movendo um passo mais perto do que se espera de um sistema artificial geral de inteligência (CARLEO et al, 2019).

Com base no problema dado e nos dados disponíveis, podemos distinguir três tipos de ML: aprendizado supervisionado, aprendizado não supervisionado e aprendizado por reforço. Enquanto muitos aplicativos em mercados eletrônicos usam aprendizado supervisionado (BRYNJOLFSSON & MCAFEE 2017), por exemplo, para prever mercados de ações (JAYANTH BALAJI et al. 2018), para entender as percepções do cliente (RAMASWAMY & DECLERCK 2018), para analisar as necessidades do cliente (KÜHL et al. al. 2020), ou para pesquisar produtos (BASTAN et al. 2020), existem implementações de todos os tipos, por exemplo, criação de mercado com aprendizado por reforço (SPOONER et al. 2018) ou segmentação de mercado não supervisionada usando avaliações de clientes (AHANI et al. al. 2019).

O aprendizado de máquina depende de diferentes algoritmos para resolver problemas de dados. O tipo de algoritmo empregado depende do problema, do número de variáveis e do modelo.

2.4.1 Naive Bayes

É uma técnica de classificação baseada no Teorema de Bayes com uma suposição de independência entre os preditores. Um classificador Naive Bayes assume que a presença de uma característica particular em uma classe não está relacionada com a presença de qualquer outro recurso. É usado principalmente para agrupamento e propósito de classificação depende da probabilidade condicional de acontecer.

Assumimos que um conjunto de dados contém n instâncias (ou casos) x_i , $i = 1..n$, que consiste em p atributos, ou seja, $x_i = (x_{i1}, x_{i2}, \dots, x_{ip})$. Cada instância é assumida como

pertencente a uma (e apenas uma) classe $y \in \{y_1, y_2, \dots, y_c\}$. A maioria dos modelos preditivos em aprendizado de máquina gera uma pontuação numérica s para cada instância x_i . Esta pontuação quantifica o grau de pertinência de classe desse caso na classe y_j . Se o conjunto de dados contiver apenas dados positivos e instâncias negativas, $y \in \{0, 1\}$, então um modelo preditivo pode ser usado como classificador.

O algoritmo Naive Bayes refere-se à construção de um modelo probabilístico Bayesiano que atribui uma probabilidade de classe para uma instância: $P(Y = y_j | X = x_i)$. O classificador simples de Naive Bayes usa essas probabilidades para atribuir uma instância a uma classe.

Uma vantagem do classificador Naive Bayes é que ele requer apenas uma pequena quantidade de dados de treinamento para estimar os parâmetros necessários para a classificação (BERRAR, 2018).

2.4.2 K Nearest Neighbor

O classificador K- Nearest Neighbor (KNN) vem na categoria de um *lazy learner*. O método consiste em armazenar a tupla de treinamento, e espera por uma tupla de teste. No KNN todas as tuplas de treinamento são armazenadas em um espaço n dimensional. Quando uma tupla dos dados de teste é fornecida ao classificador, ele busca as k tuplas de treinamento que são mais próximo da tupla desconhecida. Essas k tuplas selecionadas são k vizinho mais próximo da tupla desconhecida (dado de teste). Para classificar um registro desconhecido a distância entre outros registros de treinamento são computados. Com base na distância dos K vizinhos mais próximos são identificados e os rótulos de classe desses vizinhos mais próximos é usado para determinar o rótulo de classe de registro desconhecido.

A proximidade ou proximidade de uma tupla é definida pela métrica de distância. Uma métrica de distância comumente usada é distância euclidiana e é calculado como:

$$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Para k vizinhos mais próximos, a classe prevista da amostra de teste x é igual à classe de treino mais frequente entre k amostras de treinamento mais próximas. (AGRAWAL, 2014)

2.4.3 Support Vector Machine

Máquina de Vetores de Suporte (*Support Vector Machines*- SVMs) é um método de aprendizado relativamente novo usado para classificação binária. Tem como objetivo criar

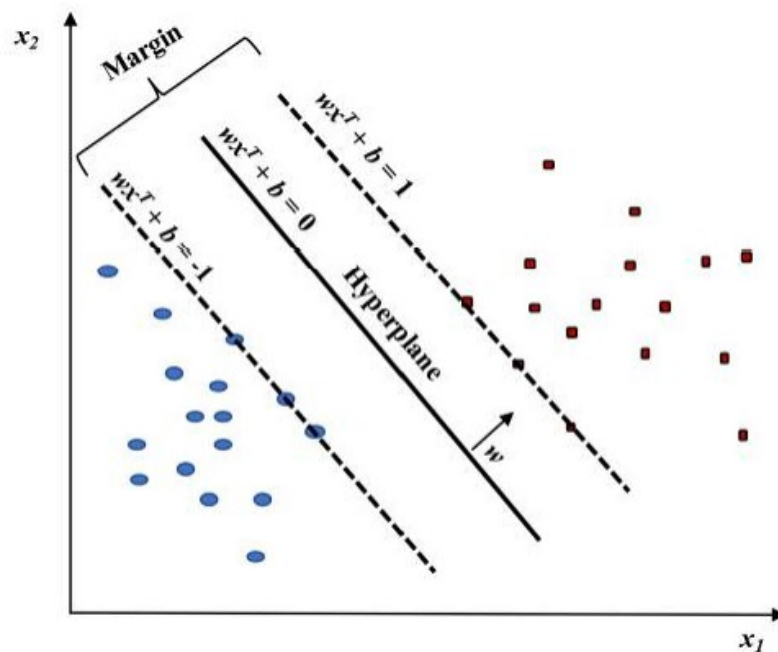
um limite de decisão entre duas classes que permite a previsão de novos valores classificando os mesmos. Este limite de decisão, conhecido como hiperplano, está orientado de tal forma que está o mais longe possível a partir dos pontos de dados mais próximos de cada uma das classes. Esses pontos mais próximos são chamados de vetores de suporte. Dado um conjunto de dados de treinamento rotulado: $(x_1, y_1), \dots, (x_n, y_n)$, $x_i \in R^d$ e $y_i \in (-1, +1)$ onde x_i é uma representação de vetor de recursos e y_i o rótulo de classe (negativo ou positivo) de um composto de treinamento i . O hiperplano ótimo pode então ser definido como: $w x^T + b = 0$ onde w é o vetor de peso, x é o vetor de característica de entrada e b é o viés (*bias*). O w e b satisfariam as seguintes desigualdades para todos os elementos do conjunto de treinamento:

$$w x_i^T + b \geq +1 \text{ se } y_i = 1$$

$$w x_i^T + b \leq -1 \text{ se } y_i = -1$$

O objetivo de treinar um modelo SVM é encontrar o w e b para que o hiperplano separe os dados e maximize a margem $\frac{1}{\|w\|^2}$. Vetores x_i para os quais $|y_i| (w x_i^T + b) = 1$ será denominado vetor de suporte (Figura 2) (HUANG et al., 2018).

Figura 2 - Modelo SVM linear. Duas classes (vermelho versus azul) foram classificadas.



Fonte: HUANG et al. (2018).

No entanto, uma vez que os dados de exemplo muitas vezes não são linearmente separáveis, os SVMs introduzem a noção de um “espaço de recursos induzido por kernel” que lança os dados em um espaço dimensional onde os dados são separáveis. No geral, os SVMs são intuitivos, teoricamente bem fundamentados e têm se mostrado praticamente bem sucedido.

SVMs também foram estendidos para resolver tarefas de regressão, onde o sistema é treinado para produzir um valor numérico (HUANG et al., 2018).

2.4.4 Regressão Logística

Os modelos de regressão constituem uma das ferramentas estatísticas mais importantes na análise estatística de dados quando se pretende modelar relações entre variáveis. O principal objetivo destes modelos é explorar a relação entre uma ou mais variáveis explicativas (ou independentes) e uma variável resposta (ou dependente). Um dos casos particulares dos modelos lineares generalizados são os modelos onde a variável resposta apresenta apenas duas categorias ou que de alguma forma dual assumindo valores 0 ou 1 sendo o modelo de regressão logística o mais popular desses modelos (CABRAL, 2013).

Métodos multivariáveis de análise estatística comumente aparecem na literatura geral de ciências da saúde (BAGLEY et al., 2001). Os termos "análise multivariada" e "análise multivariável" são frequentemente usados de forma intercambiável na literatura. Em sentido estrito, a análise multivariada refere-se à previsão simultânea de múltiplos resultados e a análise multivariável usa múltiplas variáveis para prever um único resultado (KATZ, 1999).

Os métodos multivariáveis exploram uma relação entre duas ou mais variáveis preditoras (independentes) e uma variável de resultado (dependente). O modelo que descreve a relação expressa o valor previsto da variável de resultado como uma soma de produtos, cada produto formado pela multiplicação do valor e do coeficiente da variável independente. Os coeficientes são obtidos como o melhor ajuste matemático para o modelo especificado. Um coeficiente indica o impacto de cada variável independente na variável de resultado, ajustando-se para todas as outras variáveis independentes. O modelo serve a dois propósitos: pode prever o valor da variável dependente para novos valores das variáveis independentes e pode ajudar a descrever a contribuição relativa de cada variável independente para a variável dependente, controlando as influências das outras variáveis independentes. Os quatro principais métodos multivariáveis usados em ciências da saúde são regressão linear, regressão logística, análise discriminante e regressão de risco proporcional.

Os quatro métodos multivariáveis têm muitas semelhanças matemáticas, mas diferem na expressão e no formato da variável de resultado. Na regressão linear, a variável de resultado é uma quantidade contínua, como a pressão arterial. Na regressão logística, a variável de resultado geralmente é um evento binário, como reagente versus não reagente. Na análise discriminante, a variável de resultado é uma categoria ou grupo ao qual uma entrada pertence.

Para apenas duas categorias, a análise discriminante produz resultados semelhantes à regressão logística.

A regressão logística é uma técnica estatística que tem como objetivo modelar, a partir de um conjunto de observações, a relação “logística” entre uma variável resposta dicotômica e uma série de variáveis explicativas numéricas (contínuas, discretas) e/ou categóricas (CABRAL, 2013). A regressão logística é o método multivariável mais popular usado nas ciências da saúde (TETRAULT et al., 2008).

A regressão logística, às vezes chamada de modelo logístico ou modelo logit, analisa a relação entre várias variáveis independentes e uma variável dependente categórica e estima a probabilidade de ocorrência de um evento ajustando os dados a uma curva logística. Existem dois modelos de regressão logística, regressão logística binária e regressão logística multinomial. A regressão logística binária é normalmente usada quando a variável dependente é dicotômica e as variáveis independentes são contínuas ou categóricas. Quando a variável dependente não for dicotômica e for composta por mais de duas categorias, uma regressão logística multinomial pode ser empregada.

A probabilidade de um evento é a razão entre a probabilidade de um evento ocorrer e a probabilidade de que ele não ocorra. Se a probabilidade de um evento ocorrer é p , a probabilidade de o evento não ocorrer é $(1-p)$. Então a probabilidade correspondente é um valor dado por:

$$\text{Odds of event} = \frac{p}{1-p} \quad (2)$$

Como a regressão logística calcula a probabilidade de um evento ocorrer sobre a probabilidade de um evento não ocorrer, o impacto das variáveis independentes geralmente é explicado em termos de probabilidades. Com a regressão logística a média da variável resposta p em função de uma variável explicativa x é modelada relacionando p e x através da equação $p = \alpha + \beta x$. Este não é um bom modelo pois valores extremos de x darão valores de $\alpha + \beta x$ que não caem entre 0 e 1. A solução de regressão logística para este problema é transformar as probabilidades usando o logaritmo natural (PENG et al., 2002). Com a regressão logística modelamos as probabilidades logarítmicas naturais como uma função linear da variável explicativa:

$$\text{logit}(y) = \ln \ln(\text{odds}) = \ln \ln \left(\frac{p}{1-p} \right) = \alpha + \beta x \quad (3)$$

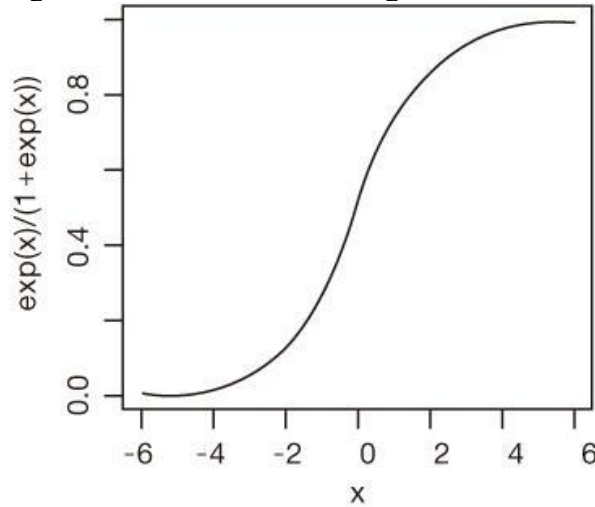
onde p é a probabilidade do resultado interessado e x é a variável independente. Os parâmetros da regressão logística são α e β . Este é o modelo logístico simples.

Tomando o antilogaritmo da equação (3) em ambos os lados, pode-se derivar uma equação para a previsão da probabilidade de ocorrência do resultado de interesse como

$$p = \frac{e^{a+\beta x}}{1 + e^{a+\beta x}} = \frac{1}{1 + e^{-(a+\beta x)}} \quad (4)$$

A regressão logística ajusta α e β , os coeficientes de regressão. A Figura 3 mostra a função logística quando α e β são 0 e 1, respectivamente.

Figura 3 - Gráfico da curva logística onde $\alpha=0$ e $\beta=1$.



Fonte: PARK (2013).

Estendendo a lógica da regressão logística simples para múltiplos preditores, pode-se construir uma regressão logística complexa como:

$$\text{logit}(y) = \ln \ln(\text{odds}) = \ln \ln \left(\frac{p}{1-p} \right) = a + \beta_1 x_1 + \dots + \beta_k x_k \quad (5)$$

portanto,

$$p = \frac{e^{a+\beta_1 x_1 + \dots + \beta_k x_k}}{1 + e^{a+\beta_1 x_1 + \dots + \beta_k x_k}} = \frac{1}{1 + e^{-(a+\beta_1 x_1 + \dots + \beta_k x_k)}} \quad (6)$$

A regressão logística é um método para ajustar uma curva de regressão, $y=f(x)$, quando y consiste em dados codificados binários. Quando a resposta é uma variável binária (dicotômica) e x é numérica, a regressão logística ajusta uma curva logística à relação entre x e y . A curva logística é uma curva em forma de S ou sigmoide, frequentemente usada para modelar o crescimento populacional (EBERHARDT & BREIWICK, 2012).

Uma função logística simples é definida pela fórmula

$$y = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-x}} \quad (7)$$

para proporcionar flexibilidade, a função logística pode ser estendida para a forma:

$$y = \frac{e^{a+\beta x}}{1 + e^{a+\beta x}} = \frac{1}{1 + e^{-(a+\beta x)}} \quad (8)$$

onde α e β determinam o intercepto e a inclinação.

3 MATERIAIS E MÉTODOS

De acordo com Bassert, J.M. et al. (2021) a história e o exame físico são os primeiros passos na observação do técnico de qualquer paciente ou grupo de pacientes. As informações obtidas nesses processos servem de base para todas as avaliações e intervenções subsequentes. É fundamental que os técnicos veterinários possam obter informações históricas completas e precisas nas avaliações de cada paciente e grupo. Da mesma forma, boas habilidades de exame físico permitem a identificação rápida de problemas significativos, seguida de medidas terapêuticas adequadas. O exame físico inclui uma avaliação profissional da saúde e bem-estar do paciente.

Neste trabalho, o banco de dados foi criado a partir de registros de exames clínicos existentes em 340 (trezentos e quarenta) cães (casos: $n = 177$, não casos: $n = 163$). Obtivemos dezessete variáveis que descrevem as características do cão, conforme apresentado na Tabela 1. Essas variáveis, segundo os veterinários, são variáveis que eles observam em uma primeira visão do animal suspeito de LV: Sexo, presença de ectoparasitas, nutrição, linfonodos, coloração da mucosa, sangramento, pelagem, lesão no focinho e/ou orelha, unhas, presença de lesão de pele, despigmentação, alopecia, secreção ocular, blefarite, proximidade da mata e o ELISA (*En-zy-me-Linked Immunosorbent Assay*) resultados do teste. Com essas informações dos veterinários, o passo inicial é utilizar essas variáveis para treinar os modelos. Na Tabela 1 também podemos ver que p = valor de um teste de correlação entre os níveis das variáveis.

Tabela 1 - Estatísticas descritivas e associações univariáveis de características incluídas na previsão de aprendizado de máquina da leishmaniose visceral canina (casos: $n = 177$; não casos: $n = 163$).

Variável	Categoria	Não Reagentes	Reagentes	<i>p-value</i>
Sexo	Fêmea	74	74	0.505
	Macho	89	103	
Idade (Meses)	Média/Desvio Padrão	34.39/30.8	44.03/36.45	0.009
Condição	Apático	19	27	0.346
	Ativo	144	150	
Presença de ectoparasitas	Não	127	151	0.078
	Sim	36	26	
Nutrição	Normal	119	113	0.058
	Magro	41	53	
	Caquético	3	11	

Linfonodos	Normal	25	27	0.983
	Aumentado	138	150	
Coloração Mucosa	Normal	121	123	0.332
	Pálida	42	54	
Sangramento	Não	156	162	0.118
	Sim	7	15	
Pelagem	Normal	87	65	0.007
	Regular	44	70	
	Ruim	32	42	
Lesão no focinho/orelha	Não	133	118	0.002
	Sim	30	177	
Unhas	Aumentadas	127	100	<0.001
	Onicogribose	36	77	
Presença de lesão na pele	Não	153	161	0.314
	Sim	10	16	
Despigmentação	Não	162	177	0.479
	Sim	1	0	
Alopecia	Não	116	89	<0.001
	Sim	47	88	
Secreção ocular	Não	159	166	0.115
	Sim	4	11	
Blefarite	Não	145	157	0.94
	Sim	18	20	
Proximidade da Mata	Não	77	141	<0.001
	Sim	86	36	

Fonte: Elaborado pelo autor (2021).

A coleta de dados foi realizada em algumas regiões do oeste do estado do Maranhão (1°59'–4° 00'S e 44° 21'–45°33' W), que é baixo no Índice de Desenvolvimento Humano (IDH) (LARIOS et al., 2021) (Figura 4). Está dividido em 21 (Vinte e um) municípios. Dos 21 municípios que compõem a Microrregião da Baixada Maranhense, 10 municípios foram utilizados na pesquisa. As cidades escolhidas foram Pinheiro, Arari, Matinha, Viana e São Bento, Bela Vista do Maranhão, Monção, Cajari, Pedro do Rosário e Conceição do Lago Açu.

Figura 4 - Região do Maranhão, Brasil, onde os dados foram coletados.



Fonte: Elaborado pelo autor (2021).

3.1 Seleção Das Variáveis E Modelo De Aprendizagem

A variável alvo, resultados do teste ELISA, é dicotômica e a Regressão Logística (LR) aparece como uma boa escolha para o algoritmo de aprendizado (KLEINBAUM et al., 2002; HENDRIX et al., 2016; NEUBER et al., 2017). Quatro algoritmos foram testados para escolher o melhor modelo: Support Vector Machine (SVM), K-Nearest Neighbor (KNN), Naive Bayes (NB) e Regressão Logística (LR), além do classificador K-nearest neighbor, que é baseado na característica do k-vizinho mais próximo de um novo ponto (amostra) para classificá-lo. Neste trabalho, os melhores resultados foram alcançados com $k = 10$.

O classificador Naive Bayes baseia-se na suposição de independência entre as variáveis do problema. O modelo NB realiza uma classificação probabilística de uma amostra não classificada para colocá-la na classe mais provável.

A máquina de vetores de suporte é um modelo de alto desempenho para problemas não lineares, não tendencioso por outliers e não sensível a eles. Inclui a Classificação do Vetor de Suporte (SVC) e Regressão do Vetor de Suporte (SVR) (XIN et al., 2018).

Para cada modelo, aplicamos uma eliminação de recursos recursivos com validação cruzada como etapa de pré-processamento (BROWNLEE et al., 2020; PEDREGOSA et al., 2011) para selecionar as melhores variáveis. Para o modelo SVM, as variáveis excluídas foram idade, linfonodos e secreção ocular. Para o modelo LR, NB e KNN as variáveis excluídas foram idade, condição e despigmentação. Os algoritmos foram treinados e testados com o conjunto de dados contendo apenas as melhores variáveis.

Os modelos de regressão são uma das ferramentas estatísticas mais importantes na análise estatística de dados para modelagem de relacionamentos entre variáveis. Esses modelos visam detectar a relação entre uma ou mais variáveis explicativas e a resposta, ou variáveis dependentes. Um dos casos particulares de modelos logísticos generalizados é aquele em que a variável resposta possui apenas duas categorias de valores dicotomizados (0 ou 1) (HOFFMAN, 2021).

A regressão logística visa modelar, a partir de um conjunto de observações, a relação logística (distribuição de probabilidade) entre uma variável resposta dicotômica e uma série de variáveis explicativas numéricas, que podem ser contínuas, discretas e categóricas (HOFFMAN, 2021; KLEINBAUM et al., 2016). A ideia é usar a expressão logística dada por:

$$y = (1 + e^{-z})^{-1} \quad (2)$$

onde $z = a_0 + aTX$, X é uma matriz $m \times n$ contendo m exemplos com n características, y é uma matriz $m \times 1$ de 0 e 1, e a é um vetor $m \times 1$ contendo os parâmetros do sistema, que ser inferido pelo algoritmo de aprendizado. Essa inferência é feita por uma tarefa iterativa com o objetivo de minimizar o erro entre os valores reais e os valores inferidos de y . Após obter o vetor de parâmetros, a , pode-se inferir um valor para uma nova amostra. A classificação é feita da seguinte forma: O algoritmo de aprendizagem irá, para cada exemplo, determinar um número pela equação (1), que representa a probabilidade de $y = 1$, e se este número for igual ou superior a 0,5 colocará $y = 1$, ou 0 caso contrário. Com o parâmetro vetor a , podemos atribuir um número para cada novo vetor de característica de cão mostrado ao sistema.

Por exemplo, podemos atribuir a matriz de características 240×14 (após a seleção das variáveis), X , contendo os valores das 14 características para cada um dos 240 cães e um vetor 240×1 , y , contendo os valores 0 ou 1 se o cão não tem ou tem a doença. Separamos essa amostra em duas partes: 80% para treinamento e 20% para teste. Para o conjunto de treinamento, apresentamos o algoritmo de aprendizado com uma matriz 192×14 e um vetor y correspondente de 192×1 . Nesta fase de aprendizado, o algoritmo estimará os parâmetros a_i , $i = 0, \dots, 14$. Com o vetor estimado a , podemos obter z e colocá-lo na Equação (2) para estimar os valores de y para cada amostra no conjunto de teste. Com os valores inferidos e reais, podemos obter a matriz de confusão para obter as métricas explicadas na próxima seção. Neste trabalho, após a fase de treinamento, temos $a = [0.00401544, -0.3851127, 0.25599153, -0.05313171, 0.54447301, 0.36488774, -0.21396936, 0.15184775, 0.28558144, -0.11643821, 0.47422156, 0.398408060, -0.34161375, -1,24370334] T$ e $a_0 = 0,30852909$.

3.2 Teste Diagnóstico

Diagnosticar uma doença é uma questão delicada porque está em jogo a vida dos pacientes, sejam eles humanos, cães ou até plantas. As ferramentas utilizadas no processo diagnóstico são testes baseados em medidas feitas em pacientes, sejam quantitativas ou qualitativas, chamados testes clínicos ou testes diagnósticos (HENDRIX et al., 2016; NEUBER et al., 2017).

Essas ferramentas se tornaram tão importantes e difundidas que existem grandes indústrias e laboratórios inteiramente dedicados à produção de testes diagnósticos cada vez mais precisos, rápidos e baratos. Os testes podem ser enganosos, especialmente onde pode haver um problema com um sistema biológico. Antes de um teste ser usado como auxiliar no diagnóstico de uma determinada doença, seu potencial de erro deve ser avaliado (HENDRIX et al., 2016; NEUBER et al., 2017).

As propostas tecnológicas para reduzir os custos financeiros do tratamento de doenças e o uso de exames laboratoriais gerais são de grande interesse. Essas tecnologias funcionam como testes de triagem deixando os exames laboratoriais para serem realizados apenas em seres com alta probabilidade de presença de doenças. As técnicas de aprendizado de máquina podem ajudar a identificar indivíduos doentes com uma probabilidade estatística razoável de verdadeiros positivos.

Yang Xin (et al., 2018) dizem: “O modelo de avaliação é uma parte muito importante da missão de aprendizado de máquina”. Neste trabalho, seguimos seus passos para avaliar nossa proposta, utilizando as métricas obtidas a partir da matriz de confusão. A matriz de confusão é mostrada na Tabela 2.

Tabela 2 - Indicações da matriz de confusão.

	Previsto como Negativo	Previsto como Positivo
Rotulado como Negativo	Verdadeiro Negativo/ <i>True Negative</i> (TN)	Falso Positivo/ <i>False Positive</i> (FP)
Rotulado como Positivo	Falso Negativo/ <i>False Negative</i> (FN)	Verdadeiro Positivo/ <i>True Positive</i> (TP)

Fonte: Elaborado pelo autor (2021).

Além disso, as seguintes métricas podem ser calculadas a partir da matriz de confusão:

- Precisão: $(TN + TP) / (TN + FP + FN + TP)$. Isso mede a fração de previsões corretas.

- Sensibilidade ou Recall: $(TP) / (TP + FN)$. Isso mede a capacidade do teste para identificar corretamente os indivíduos que têm a doença. Ele mede a probabilidade de o teste obter um resultado positivo, desde que a condição verdadeira esteja presente. Essa é a métrica mais importante na triagem, pois um resultado negativo em um teste com alta sensibilidade é útil para excluir a existência da condição.
- Especificidade: $TN / (TN + FP)$. Essa é a capacidade do teste em identificar corretamente indivíduos que não possuem a doença;
- O Valor Preditivo Positivo ou Precisão: $TP / (TP + FP)$. Isso mede a probabilidade de o cão ter a doença sabendo que o resultado do teste é positivo.
- O Valor Preditivo Negativo: $TN / (TN + FN)$. Isso mede a probabilidade de o cão não ter a doença sabendo que o resultado do teste é negativo.
- A Razão de Verossimilhança Positiva (RV+): Sensibilidade / (1-Especificidade). Isso mostra que para um valor maior que 1 (um), o teste positivo é mais provável de ocorrer em cães com a doença do que naqueles sem a doença;
- A Razão de Verossimilhança Negativa (RV-): (1-Sensibilidade) /Especificidade. Isso mostra que para um valor menor que 1 (um), o teste negativo é mais provável de ocorrer em cães com a doença do que naqueles sem a doença.
- Área Sob Curva (AUC): a partir da curva Receiver Operating Characteristic (ROC): Isso é realizado para identificar quão bom o modelo desenvolvido é para distinguir entre dois parâmetros, a taxa de verdadeiro positivo e a taxa de falso negativo. Modelos com previsões 100% corretas têm uma AUC de 1.

4 RESULTADOS E DISCUSSÕES

O banco de dados foi dividido em oitenta por cento para treinamento e vinte por cento para teste. O conjunto de treinamento foi utilizado para determinar os parâmetros do algoritmo de aprendizado e o conjunto de teste foi utilizado para validar os modelos pelas métricas descritas acima. A Tabela 3 mostra os resultados para os modelos utilizados.

Tabela 3 - Métricas de teste dos modelos. Pode-se ver que a Regressão Logística obteve as melhores métricas.

	Naive Bayes (NB)	K Nearest Neighbor (KNN)	Support Vector Machine (SVM)	Logistic Regression (LR)
Acurácia	0.63	0.63	0.69	0.75

Sensibilidade (Recall)	0.56	0.56	0.84	0.84
Especificidade	0.69	0.69	0.56	0.67
Valor Preditivo Positivo	0.62	0.62	0.63	0.69
Valor preditivo negativo	0.64	0.64	0.80	0.83
RV+	1.84	1.84	1.90	2.53
RV-	0.63	0.63	0.28	0.23
AUROC (<i>Area Under Curve</i>)	0.71	0.71	0.70	0.77

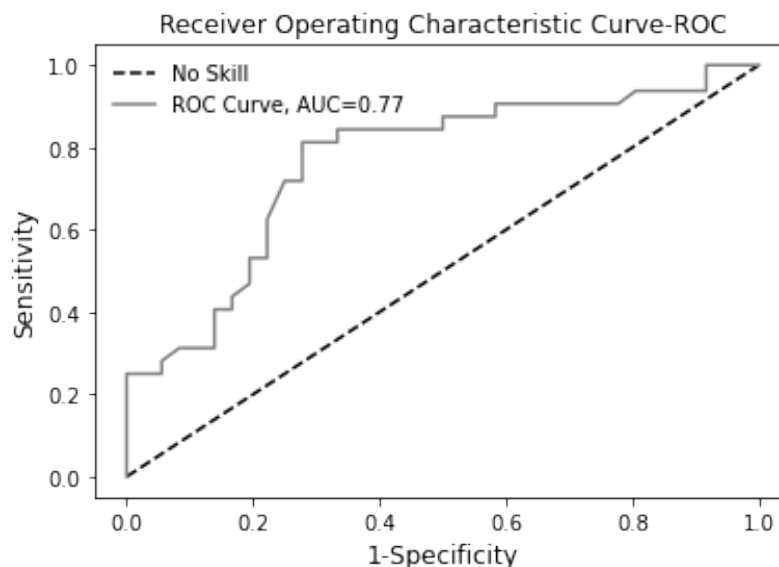
Fonte: Elaborado pelo autor (2021).

A partir da Tabela 3, pode-se observar que o modelo LR obteve melhores resultados que os demais, por exemplo, acurácia de 75%, sensibilidade de 84% e valor preditivo negativo de 83%. Estes são bons resultados porque nos dão uma grande segurança de que o cão previsto como não tendo a doença, na verdade não a tem; então não é necessário realizar um exame laboratorial nele. Um valor preditivo positivo de 0,69 significa que temos aproximadamente 70% de certeza de que os cães testados como positivos realmente têm a doença. O RV+ igual a 2,53 significa que um cão com a doença tem 2,53 vezes mais chances de ter um teste positivo do que um sem a doença. O RV- igual a 0,23 significa que um cão sem a doença tem, aproximadamente, quatro (0,25) vezes mais chance de testar negativo do que aqueles com a doença. Assim, o modelo de regressão logística mostra uma boa capacidade em rejeitar falsos negativos.

A AUC de 0,77 (Figura 5) mostra a capacidade discriminatória do teste para distinguir entre cães com e sem a doença.

A partir desses resultados, pode-se observar que o modelo de regressão logística pode atuar como um método de triagem eficiente para cães com leishmaniose visceral canina com base apenas na sua visualização e, assim, reduzir o custo em exames laboratoriais.

Figura 5 - Curva ROC para aplicação do modelo LR no conjunto de testes.



Fonte: Elaborado pelo autor (2021).

Na tentativa de compreender o tipo de amostras classificadas corretamente e não corretamente, para o modelo de Regressão Logística, obtemos as características descritivas de cada variável para as quatro classificações: TN, FP, FN e TP.

Na Tabela 4 temos a matriz de confusão. Pode-se ver que este modelo obteve 5 falsos negativos e 12 falsos positivos.

Tabela 4 - Matriz de confusão.

	Previsto como Negativo	Previsto como Positivo
Rotulado como Negativo	24	12
Rotulado como Positivo	5	27

Fonte: Elaborado pelo autor (2021).

Na Tabela 5 observa-se que as amostras classificadas como falso-negativas. Pode-se observar que 100% das amostras apresentam as seguintes características: coloração normal da mucosa, ausência de sangramento, unhas aumentadas, ausência de lesão de pele, ausência de secreção ocular e ausência de blefarite.

Tabela 5 - Frequência das variáveis para as amostras de falsos negativos.

Variável	Categoria	Frequência (%)
Sexo	Fêmea	20
	Macho	80
Presença de ectoparasitas	Não	80
	Sim	20
Nutrição	Normal	60
	Magro	40
	Caquetico	0
Linfonodos	Normal	20
	Aumentado	80
Cor da mucosa	Normal	100

	Pálido	0
Sangramento	Não	100
	Sim	0
Pelagem	Normal	20
	Regular	60
	Ruim	20
Lesão no focinho/orelha	Não	80
	Sim	20
Unhas	Aumentada	100
	Onicogrífose	0
Presença de lesão na pele	Não	100
	Sim	0
Alopecia	Não	80
	Sim	20
Secreção ocular	Não	100
	Sim	0
Blefarite	Não	100
	Sim	0
Proximidade da mata	Não	20
	Sim	80

Fonte: Elaborado pelo autor (2021).

Na Tabela 6 observa-se que as amostras classificadas como falso positivo. Pode-se observar que 100% das amostras apresentam as seguintes características: Sem sangramento e sem secreção ocular.

Tabela 6 - Frequência das variáveis para as amostras falso-positivas.

Variável	Categoria	Frequência (%)
Sexo	Fêmea	58
	Macho	42
Presença de ectoparasitas	Não	0.92
	Sim	0.08
Nutrição	Normal	83
	Magro	17
	Caquetico	0
Linfonodos	Normal	8
	Aumentado	92
Cor da mucosa	Normal	75
	Pálido	25
Sangramento	Não	100
	Sim	0
Pelagem	Normal	50
	Regular	30
	Ruim	20
Lesão no focinho/orelha	Não	83
	Sim	17
Unhas	Aumentada	58
	Onicogrífose	42
Presença de lesão na pele	Não	83

	Sim	17
	Não	42
Alopecia	Sim	58
	Não	100
Secreção ocular	Sim	0
	Não	92
Blefarite	Sim	8
	Não	92
Proximidade da mata	Sim	8

Fonte: Elaborado pelo autor (2021).

A Tabela 7 mostra as amostras classificadas como verdadeiro negativo. Pode-se observar que as amostras apresentam as seguintes características: Sem presença de lesão de pele e sem secreção ocular.

Tabela 7 - Frequência das variáveis para as amostras de verdadeiros negativos.

Variável	Categoria	Frequência (%)
Sexo	Fêmea	42
	Macho	58
Presença de ectoparasitas	Não	75
	Sim	25
Nutrição	Normal	83
	Magro	17
	Caquetico	0
Linfonodos	Normal	8
	Aumentado	92
Cor da mucosa	Normal	67
	Pálido	33
Sangramento	Não	92
	Sim	8
Pelagem	Normal	62
	Regular	21
	Ruim	17
Lesão no focinho/orelha	Não	96
	Sim	4
Unhas	Aumentada	96
	Onicogribose	4
Presença de lesão na pele	Não	0
	Sim	100
Alopecia	Não	92
	Sim	8
Secreção ocular	Não	100
	Sim	0
Blefarite	Não	96
	Sim	4
Proximidade da mata	Não	21
	Sim	79

Fonte: Elaborado pelo autor (2021).

A Tabela 8 mostra as amostras classificadas como verdadeiras positivas. Pode-se observar que a maioria das amostras verdadeiras positivas não tem presença de ectoparasitas, linfonodos aumentados, sem presença de sangramento, sem presença de lesão de pele, sem secreção ocular, sem blefarite e sem proximidade com a floresta.

Tabela 8 - Frequência das variáveis para as amostras verdadeiras positivas.

Variável	Categoria	Frequência (%)
Sexo	Fêmea	30
	Macho	70
Presença de ectoparasitas	Não	93
	Sim	7
Nutrição	Normal	67
	Magro	22
	Caquetico	11
Linfonodos	Normal	12
	Aumentado	88
Cor da mucosa	Normal	63
	Pálido	37
Sangramento	Não	89
	Sim	11
Pelagem	Normal	26
	Regular	37
	Ruim	37
Lesão no focinho/orelha	Não	63
	Sim	37
Unhas	Aumentada	44
	Onicogribose	56
Presença de lesão na pele	Não	85
	Sim	15
Alopecia	Não	40
	Sim	60
Secreção ocular	Não	85
	Sim	15
Blefarite	Não	93
	Sim	7
Proximidade da mata	Não	97
	Sim	3

Fonte: Elaborado pelo autor (2021).

5 CONSIDERAÇÕES FINAIS

Neste trabalho, quatro modelos de aprendizado de máquina foram testados como método inicial no atendimento veterinário para identificar cães com leishmaniose visceral canina com base apenas na inspeção visual e física do animal. Para isso, obtivemos dados clínicos de 340 cães com dezoito variáveis. Essas variáveis foram escolhidas com base nas experiências dos profissionais veterinários e para cada modelo foram selecionadas as melhores

variáveis para prever os resultados. Os modelos testados foram regressão logística, máquina de vetores de suporte, vizinho K-mais próximo e Naïve Bayes. O modelo de regressão logística, utilizando quatorze variáveis após o procedimento de seleção de variáveis, obteve as melhores métricas: acurácia de 75%, sensibilidade de 84%, especificidade de 67%, razão de verossimilhança positiva de 2,53 e razão de verossimilhança negativa de 0,23.

Observa-se então que o modelo foi capaz de identificar a doença com certo grau de certeza nos animais, esse modelo possibilita a redução de custos nesse tipo de atendimento, com o objetivo de realizar a triagem do animais em áreas endêmicas possibilitando aos profissionais da área da saúde tomarem decisões com base nas informações obtidas, sendo possível por exemplo, separar os animais suspeitos dos outros que não apresentam a doença e realizar exame laboratorial posterior nos animais triados, dessa forma contribuindo para a melhoria da saúde pública urbana.

O presente trabalho também deu origem a publicação no periódico científico *MDPI Sensors*, de excelência internacional, com o título *Diagnostic Classification of Cases of Canine Leishmaniasis Using Machine Learning* (FERREIRA et. al., 2022). Mostrando-se de interesse de toda comunidade científica, possibilitando levantar questionamentos e inspirar trabalhos futuros na presente área tão importante para todo o mundo, para o Brasil e o Maranhão.

REFERÊNCIAS

- ABBIATI, Thaís Carneiro et al. Leishmaniose visceral canina: Relato de caso. **Pubvet**, v. 13, p. 152, 2019.
- ABRANCHES, P.; CAMPINO, L.; SANTOS-GOMES, G. M. Canine leishmaniasis. New concepts of epidemiology and immunopathology: their impact in the control of human visceral leishmaniasis. **Acta médica portuguesa**, v. 11, n. 10, p. 871-5, 1998.
- ADLER, S.; THEODOR, O. Investigations on Mediterranean kala-azar. II — *Leishmania infantum*. **Proceedings of Royal Society of London**, 108: 453-502, 1931.
- AGRAWAL, Rashmi. K-nearest neighbor for uncertain data. **International Journal of Computer Applications**, v. 105, n. 11, 2014.
- AHANI, Ali et al. Market segmentation and travel choice prediction in Spa hotels through TripAdvisor's online reviews. **International Journal of Hospitality Management**, v. 80, p. 52-77, 2019.
- ALHAJJ, Mandy; FARHANA, Aisha. Enzyme linked immunosorbent assay. **StatPearls [Internet]**, 2021.
- ALVAR, J. et al. Canine leishmaniasis: clinical, parasitological and entomological follow-up after chemotherapy. **Annals of Tropical Medicine & Parasitology**, v. 88, n. 4, p. 371-378, 1994.
- ALVAR, Jorge et al. Canine leishmaniasis. **Advances in parasitology**, v. 57, n. 3, p. 1-88, 2004.
- ALVES, Waneska Alexandra. Leishmaniose visceral americana: situação atual no Brasil Leishmaniasis: current situation in Brazil. **World Health**, v. 6, p. 25-29, 2009.
- ANDRADE, A. M.; QUEIROZ, L. H.; NUNES, G. R.; PERRI, S. H. V.; NUNES, C. M. Reposição de cães em área endêmica para leishmaniose visceral. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 40, n. 5, p. 594-595, 2007.
- ASHFORD, D.A.; DAVID, J.R.; FREIRE, M.; DAVID, R.; SHERLOCK, I.; EULÁLIO, M.C.; SAMPAIO, D.P.; BADARÓ, R. Studies on control of visceral leishmaniasis: impact of dogs control on canine and human visceral leishmaniasis in Jacobina, Bahia, Brazil. **American Journal of Tropical Medicine and Hygiene**, 59: 53-57, 1998.

BAGLEY, Steven C.; WHITE, Halbert; GOLOMB, Beatrice A. Logistic regression in the medical literature: Standards for use and reporting, with particular attention to one medical domain. **Journal of clinical epidemiology**, v. 54, n. 10, p. 979-985, 2001.

BALAJI, A. Jayanth; RAM, DS Harish; NAIR, Binoy B. Applicability of deep learning models for stock price forecasting an empirical study on BANKEX data. **Procedia computer science**, v. 143, p. 947-953, 2018.

BANETH, Gad; SHAW, Susan E. Chemotherapy of canine leishmaniosis. **Veterinary parasitology**, v. 106, n. 4, p. 315-324, 2002.

BANETH, G.; KOUTINAS, A. F.; SOLANO-GALLEGU, L.; BOURDEAU, P.; FERRER, L. Canine leishmaniosis – new concepts and insights on an expanding zoonosis: part one. **Trends in Parasitology**, v. 24, n. 7, p. 324-330, 2008.

BANETH, G.; SOLANO-GALLEGU, L. Leishmaniasis. In: GREENE, C. E. Infectious diseases of the dog and cat. 4. ed. Philadelphia: **Elsevier Saunders**, 2012. p. 735-748.

BASSERT, Joanna M. McCurnin's Clinical Textbook for Veterinary Technicians and Nurses E-Book. **Elsevier Health Sciences**, 2021.

BASTAN, M.; RAMISA, A.; TEK, M. Cross-modal fashion product search with transformer-based Embeddings. **CVPR Worksh**, v. 1, 2020.

BERRAR, Daniel. Bayes' theorem and naive Bayes classifier. **Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics**, v. 403, 2018.

BISHOP, Christopher M.; NASRABADI, Nasser M. Pattern recognition and machine learning. **New York: springer**, 2006.

BRASIL. Ministério da Saúde. Manual de vigilância e controle da Leishmaniose Visceral: normas e manuais técnicos. **Brasília: Ministério da Saúde**, 2003. 120p.

BRITO, F.L.C.; ALVES, L.C.; ORTIZ, J.P.D.; MAIA, F.C.L.; SILVA JUNIOR, V.A.; LAUS, J.L. Uveitis associated to the infection by *Leishmania chagasi* in dog from Olinda city, Pernambuco, Brazil. **Ciencia Rural**, 34: 925-929, 2004.

BROWNLEE, Jason. Data preparation for machine learning: data cleaning, feature selection, and data transforms in Python. **Machine Learning Mastery**, 2020.

BRYNJOLFSSON, Erik; MCAFEE, ANDREW. Artificial intelligence, for real. **Harvard business review**, v. 1, p. 1-31, 2017.

CABRAL, Cleidy Isolete Silva. **Aplicação do modelo de regressão logística num estudo de mercado**. 2013. Tese de Doutorado.

CARLEO, Giuseppe et al. Machine learning and the physical sciences. **Reviews of Modern Physics**, v. 91, n. 4, p. 045002, 2019.

CATÃO, Rafael de Castro. Dengue no Brasil: abordagem geográfica na escala nacional. São Paulo: Cultura Acadêmica, 2012. (**Coleção PROPG Digital - UNESP**). ISBN 9788579833328. Disponível em: <<http://hdl.handle.net/11449/109225>>.

CHAGAS, E. Primeira verificação em indivíduo vivo da Leishmaniose visceral no Brasil. **Brasil Médico**, 50: 221-222, 1936.

CHAGAS, E.; CUNHA, A.M.; FERREIRA, L.C.; DEANE, L.; DEANE, G.; GUIMARÃES, F.N.; PAUMGARTEN, M.J.; SÁ, B. Leishmaniose visceral americana (relatório dos trabalhos realizados pela comissão encarregada do estudo da leishmaniose visceral americana em 1937). **Memórias do Instituto Oswaldo Cruz**, 33: 189-229, 1938.

CHEN, Serena H.; JAKEMAN, Anthony J.; NORTON, John P. Artificial intelligence techniques: an introduction to their use for modelling environmental systems. **Mathematics and computers in simulation**, v. 78, n. 2-3, p. 379-400, 2008.

GS, Cruz. Leishmaniose tegumentar americana: aspectos clínicos, epidemiológicos e influência de fatores predisponentes. Monografia) **Universidade da Integração Internacional da Lusofonia Afro Brasileira**. Ceará, 2016.

CIARAMELLA, P.; OLIVA, G.; LUNA, R.; AMBROSIO, R.; CORTESE, L.; PERSECHINO, A.; GRADONI, L.; SCALONI, A. A retrospective clinical study of canine leishmaniasis in 150 dogs naturally infected by *Leishmania infantum*. **The Veterinary Record**, 141: 539-543, 1997.

COSTA, C.H.N. Are dogs important reservoirs of visceral leishmaniasis in Brazil? **Revista da Sociedade Brasileira de Medicina Tropical**, 30: 155, 1997.

COSTA, C.H.N.; GOMES, R.B.B.; SILVA, M.R.B.; GARCEZ, L.M.; RAMOS, P.K.S.; SANTOS, R.S.; SHAW, J.J.; DAVID, J.R.; MIGUIRE, J.H. Competence of the human host as a reservoir for *Leishmania chagasi*. **Journal of Infectious Disease**, 182: 997- 1000, 2000.

DANTAS-TORRES, Filipe; DE BRITO, Maria Edileuza Felinto; BRANDÃO-FILHO, Sinval Pinto. Seroepidemiological survey on canine leishmaniasis among dogs from an urban area of Brazil. **Veterinary parasitology**, v. 140, n. 1-2, p. 54-60, 2006.

DANTAS-TORRES, Filipe et al. *Leishmania*-FAST15: a rapid, sensitive and low-cost real-time PCR assay for the detection of *Leishmania infantum* and *Leishmania braziliensis* kinetoplast DNA in canine blood samples. **Molecular and cellular probes**, v. 31, p. 65-69, 2017.

DE LIMA JÚNIOR, Francisco Edilson Ferreira. **Cenários da leishmaniose visceral no Brasil**. 2018.

DIETZE, R.; CARVALHO, S.F.G. Leishmaniose visceral – calazar. In: CEMERMAN, S. CIMERMAN, B. (Ed). **Medicina tropical**. Ed. Atheneu, São Paulo, 2003. p. 65-84.

Indicadores socioambientais dos municípios da Baixada Maranhense são mapeados em estudo. **FAPEMA**, 2016. Disponível em: <<https://www.educamaisbrasil.com.br/enem/antropologia/principais-caracteristicas-da-cultura-afro-brasileira>>. Acesso em: 09 de nov. de 2021.

EBERHARDT, L. L.; BREIWICK, J. M. Models for population growth curves. **International Scholarly Research Notices**, v. 2012, 2012.

FARIA, Angélica Rosa; ANDRADE, Héli da Monteiro de. Diagnosis of canine visceral leishmaniasis: major technological advances and few practical applications/Diagnóstico da Leishmaniose Visceral Canina: grandes avanços tecnológicos e baixa aplicação prática. **Rev Pan-Amaz Saude**, p. 47-57, 2012.

FERRER, L.M. Clinical aspects of canine leishmaniasis. From canine leishmaniasis update (Ed. R. Killick-Kendrick). **Proceedings of a canine leishmaniasis forum**, Barcelona, 28-31 January, pp 6-10, 1999.

FERREIRA, Tiago S. et al. Diagnostic Classification of Cases of Canine Leishmaniasis Using Machine Learning. **Sensors**, v. 22, n. 9, p. 3128, 2022.

FONSECA, T. H. S. et al. Chemiluminescent ELISA with multi-epitope proteins to improve the diagnosis of canine visceral leishmaniasis. **The Veterinary Journal**, v. 253, p. 105387, 2019.

FURTADO, Aline Santos et al. Space-time analysis of visceral leishmaniasis in the State of Maranhão, Brazil. **Ciência & Saúde Coletiva**, v. 20, p. 3935-3942, 2015.

FUNASA. **Boletim Epidemiológico**. Evolução temporal das doenças de notificação compulsória no Brasil de 1980 a 1998. 1999.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. Deep learning. **MIT press**, 2016.

GOMES, Y. M. et al. Diagnosis of canine visceral leishmaniasis: biotechnological advances. **The Veterinary Journal**, v. 175, n. 1, p. 45-52, 2008.

GONTIJO, C.M.F.; MELO, M.N. Leishmaniose visceral no Brasil: quadro atual, desafios e perspectivas. **Revista brasileira de epidemiologia**, 7: 338-349, 2004.

HENDRIX, Charles M.; ROBINSON, E. D. Diagnostic parasitology for veterinary technicians-E-book. **Elsevier Health Sciences**, 2016.

HOFFMANN, John P. Linear Regression Models: Applications in R. **Chapman and Hall/CRC**, 2021.

HUANG, Shujun et al. Applications of support vector machine (SVM) learning in cancer genomics. **Cancer genomics & proteomics**, v. 15, n. 1, p. 41-51, 2018.

JORDAN, Michael I.; MITCHELL, Tom M. Machine learning: Trends, perspectives, and prospects. **Science**, v. 349, n. 6245, p. 255-260, 2015.

KATZ, Mitchell H. Multivariable analysis: a practical guide for clinicians and public health researchers. **Cambridge university press**, 2011.

KLEINBAUM, David G. et al. Logistic regression. **New York: Springer-Verlag**, 2002.

KRAUSPENHAR, C.; BECK, C.; SPEROTTO, V.; SILVA, A.A.; BASTOS, R.; RODRIGUES, L. Leishmaniose visceral em um canino de Cruz Alta, Rio Grande do Sul, Brasil. **Ciência Rural**, 37, 2007.

KÜHL, Niklas; MÜHLTHALER, Marius; GOUTIER, Marc. Supporting customer-oriented marketing with artificial intelligence: automatically quantifying customer needs from social media. **Electronic Markets**, v. 30, n. 2, p. 351-367, 2020.

LANGONI, H.; LUCHEIS, S.B.; SILVA, R.C.; CASTRO, A.P.B.; PAES, A.C. American visceral leishmaniasis: a case report. **The Journal of Venomous animals and toxins including tropical diseases**, 11: 361-372, 2005.

LARIOS, Gustavo et al. A new strategy for canine visceral leishmaniasis diagnosis based on FTIR spectroscopy and machine learning. **Journal of Biophotonics**, v. 14, n. 11, p. e202100141, 2021.

LAURENTI, M. D.; ROSSI, C. N.; MATTA, V. L. R.; TOMOKANE, T. Y.; CORBETT, C. E. P.; SECUNDINO, N. F. C.; PIMENTA, P. F. P.; MARCONDES, M. Asymptomatic dogs are highly competent to transmit *Leishmania (Leishmania) infantum* chagasi to the natural vector. **Veterinary Parasitology**, v. 196, n. 3-4, p. 296-300, 2013

LEISHMAN, W.B. On the possibility of the occurs of trypanosomiasis in Índia. **British Medical Journal**, 1: 1252-1254, 1903.

LIMA, W.G.; MICHALICK, M.S.M.; MELO, M.N.; TAFURI, W.L.; TAFURI, W.L. Canine visceral leishmaniasis: a histopathological study of lymph nodes. **Acta tropica**, 92: 43-53, 2004.

LINHARES, G.F.C.; CHAVES, N.S.T.; DUARTE, S.C.; FERNANDES, P.R.; AMARAL, A.V.C.; SOUZA, M.A. Relato de um caso clínico de leishmaniose visceral em um cão na cidade de Goiânia. **Revista de Patologia Tropical**, 34: 69-72, 2005.

MICHALICK, M.S.M.; GENARO, O. Leishmaniose Visceral Americana. In: NEVES, D.P.; MELO, A.L.; LINARDI, P.M.; VITOR, R.W.A. (Ed) *Parasitologia humana*. 11^o ed., Ed. **Atheneu**, São Paulo, 2005. p. 56-72

MIGONE, L. E. Un caso de Kalazar en Assunción (Paraguay). **Bulletin Societe Pathologic Exotique**, 6: 118-120, 1913.

MOLINA, R. et al. Infectivity of dogs naturally infected with *Leishmania infantum* to colonized *Phlebotomus perniciosus*. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 88, n. 4, p. 491-493, 1994.

NASCIMENTO, Maria do Desterro Soares Brandão et al. Prevalence of infection by *Leishmania chagasi* using ELISA (rK39 and CRUDE) and the Montenegro skin test in an endemic leishmaniasis area of Maranhão, Brazil. **Cadernos de saúde pública**, v. 21, p. 1801-1807, 2005.

NEUBER, Ariane; NUTTALL, Tim. Diagnostic techniques in veterinary dermatology. **John Wiley & Sons**, 2017.

NEVES, D.P. *Parasitologia Humana*, 13^o ed.; **Editora Atheneu: São Paulo, Brazil**, 2016

NICOLLE C., COMTE C. Origine du Kalazar. **Academy of Science**, 146: 789, 1908.

PALTRINIERI, S.; SOLANO-GALLEGO, L.; FONDATI, A.; LUBAS, G.; GRADONI, L.; CASTAGNO, M.; CROTTI, A.; MAROLI, M.; OLIVA, G.; ROURA, X.; ZATELLI, A.; ZINI, E. Guidelines for diagnosis and clinical classification of leishmaniasis in dogs. **Journal of the American Veterinary Medical Association**, v. 36, n. 11, p. 1184-1191, 2010.

PARK, Hyeoun-Ae. An introduction to logistic regression: from basic concepts to interpretation with particular attention to nursing domain. **Journal of Korean Academy of Nursing**, v. 43, n. 2, p. 154-164, 2013.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, v. 12, p. 2825-2830, 2011.

PENG, Chao-Ying Joanne; LEE, Kuk Lida; INGERSOLL, Gary M. **An introduction to logistic regression analysis and reporting**. The journal of educational research, v. 96, n. 1, p. 3-14, 2002.

PENNA, H. A. Leishmaniose visceral no Brasil. **Brasil Médico**, 48: 949-950, 1934.

PESSÔA, S.B.; MARTINS, A.V. Parasitologia médica. 11a ed. **Rio de Janeiro: Guanabara Koogan**, 1988. p. 872.

WORLD HEALTH ORGANIZATION et al. Programa das Nações Unidas para o Desenvolvimento. Atlas do desenvolvimento humano no Brasil. **World Health Organization**, 2013.

DE QUEIROZ, Nina MGP et al. Canine Visceral Leishmaniasis diagnosis by immunohistochemistry and PCR in skin tissues in association with RIFI and ELISA-test. **Revista Brasileira de Parasitologia Veterinária**, v. 19, n. 1, p. 32-38, 2010.

RAMASWAMY, Sridhar; DECLERCK, Natalie. Customer perception analysis using deep learning and NLP. **Procedia Computer Science**, v. 140, p. 170-178, 2018.

REAGAN, K. L.; REAGAN, B. A.; GILOR, C. Machine learning algorithm as a diagnostic tool for hypoadrenocorticism in dogs. **Domestic Animal Endocrinology**, v. 72, p. 106396, 2020.

REY, L. Bases da Parasitologia Médica, 2a ed. Ed. **Guanabara Koogan, Rio de Janeiro**, 2001. p. 349p

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence: A Modern Approach, eBook**. Pearson Higher Ed, 2021.

SCHOFIELD, Imogen et al. Machine-learning based prediction of Cushing's syndrome in dogs attending UK primary-care veterinary practice. **Scientific Reports**, v. 11, n. 1, p. 1-12, 2021.

SHRESTHA, Yash Raj; KRISHNA, Vaibhav; VON KROGH, Georg. Augmenting organizational decision-making with deep learning algorithms: Principles, promises, and challenges. **Journal of Business Research**, v. 123, p. 588-603, 2021.

SILVA, Francinaldo S. Patologia e patogênese da leishmaniose visceral canina. **R Trop Ci Agr Biol**, v. 1, n. 1, p. 20-31, 2007.

SIQUEIRA, Sabrina da Costa Fernandes. INSTITUTO DE SAÚDE COLETIVA. 2009. **Tese de Doutorado**. UNIVERSIDADE FEDERAL DE MATO GROSSO.

SPOONER, Thomas et al. Market making via reinforcement learning. **arXiv preprint arXiv:1804.04216**, 2018.

TETRAULT, Jeanette M. et al. Reporting of multivariable methods in the medical literature. **Journal of Investigative Medicine**, v. 56, n. 7, p. 954-957, 2008.

TORRECILHA, Rafaela Beatriz Pintor et al. Prediction of lymph node parasite load from clinical data in dogs with leishmaniasis: An application of radial basis artificial neural networks. **Veterinary parasitology**, v. 234, p. 13-18, 2017.

WERNECK, Guilherme L. et al. The urban spread of visceral leishmaniasis: clues from spatial analysis. **Epidemiology**, p. 364-367, 2002.

WORLD HEALTH ORGANIZATION et al. Control of the leishmaniasis: report of a meeting of the WHO Expert Committee on the Control of Leishmaniasis, Geneva, 22-26 March 2010. **World Health Organization**, 2010.

XIN, Yang et al. Machine learning and deep learning methods for cybersecurity. **Ieee access**, v. 6, p. 35365-35381, 2018.

