

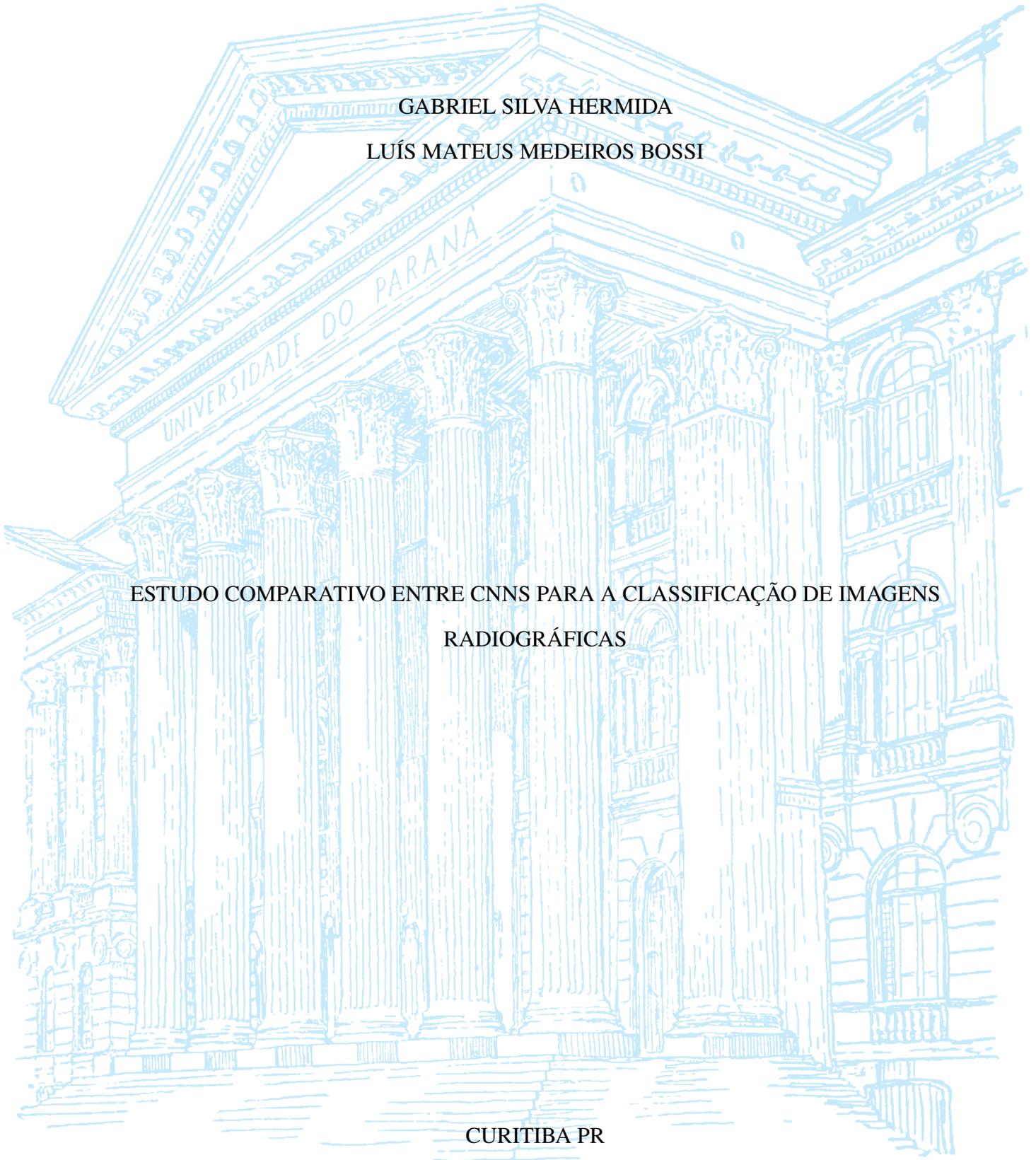
UNIVERSIDADE FEDERAL DO PARANÁ

GABRIEL SILVA HERMIDA  
LUÍS MATEUS MEDEIROS BOSSI

ESTUDO COMPARATIVO ENTRE CNNs PARA A CLASSIFICAÇÃO DE IMAGENS  
RADIOGRÁFICAS

CURITIBA PR

2023



GABRIEL SILVA HERMIDA  
LUÍS MATEUS MEDEIROS BOSSI

ESTUDO COMPARATIVO ENTRE CNNs PARA A CLASSIFICAÇÃO DE IMAGENS  
RADIOGRÁFICAS

Trabalho apresentado ao Departamento de Informática,  
pertencente ao Setor de Exatas da Universidade Federal  
do Paraná, como requisito parcial para a conclusão do  
curso Bacharelado em Ciência da Computação.

Área de concentração: *Ciência da Computação*.

Orientador: Lucas Ferrari de Oliveira.

CURITIBA PR

2023

## RESUMO

A literatura veterinária reporta que cerca de 10% dos caninos submetidos a exames de rotina são diagnosticados com problemas cardíacos, e um sintoma representativo para esses casos é o aumento do átrio esquerdo do animal. A identificação desse sintoma ocorre por meio da análise de imagens radiográficas torácicas dos animais, o que é um processo custoso para o médico veterinário. A alta incidência do sintoma, em conjunto com o custo para o diagnóstico dele, sugerem a necessidade da elaboração de uma ferramenta computacional de auxílio diagnóstico que contribua com o aumento da eficiência e precisão dos resultados, visando a redução do custo no processo de identificação do sintoma. Sob essa perspectiva, este trabalho propõe a comparação entre quatro modelos de aprendizado profundo com variações nas taxas de aprendizado de cada um deles, para a classificação das imagens entre as classes normal e com aumento de átrio. Os modelos escolhidos foram VGG16, VGG19, InceptionV3 e ResNet50V2. Para a avaliação dos resultados, obtidos através da técnica de validação cruzada, foram considerados os valores de especificidade e sensibilidade das redes. Pode-se concluir que aquela com melhor desempenho foi a InceptionV3, com valores superiores às demais em todas as métricas avaliadas cujos valores se apresentam inteiramente acima de 80%. A comparação entre essas redes específicas, bem como os resultados obtidos, representam uma contribuição importante para a área de auxílio diagnóstico, a despeito de o estudo ter algumas limitações como inconstância nos valores da *loss* nos resultados de validação da rede e a necessidade de experimentação com *datasets* maiores utilizando outras técnicas para treino, teste e validação.

Palavras-chave: Redes Neurais. Diagnóstico Assistido por Computador. Radiografia. Redes Neurais Convolucionais.

## **ABSTRACT**

The veterinary literature reports that about 10% of dogs submitted to routine examinations are diagnosed with heart diseases, and a representative symptom for these cases is the enlargement of the animal's left atrium. The identification of the anomaly is possible with the analysis of images of chest radiographs of the animals, which is a costly process for the veterinarian. The high incidence of the symptom and the high cost of diagnosing it suggests the need to develop a computational tool to aid diagnosis that contributes to increase the efficiency and accuracy of the results, aiming at reducing the cost of the diagnosis process. Under this perspective, this work proposes the comparison between four models of deep learning models with variations in the learning rates of each one of them for the classification of the images between normal classes and atrial enlargement classes. The chosen models were VGG16, VGG19, InceptionV3 and ResNet50V2, in which for the evaluation of the results, obtained through the cross-validation technique, the specificity and sensitivity values of the networks results were strongly considered. We conclude that the one with the best performance was InceptionV3. The comparison between these specific networks, as well as the results obtained, represent an important contribution to the area of computer aided diagnosis, although there are some limitations, such as inconstancy in the loss values in the network results and the need for experimentation with larger datasets using other techniques for training, testing and validation.

**Keywords:** Neural Networks. Computer Aided Diagnosis. Radiography. Convolutional Neural Networks.

## LISTA DE FIGURAS

2.1	Representação da topologia de uma rede neural artificial. . . . .	10
2.2	Representação do neurônio de uma rede neural artificial. . . . .	10
2.3	Representação visual das camadas que constituem a rede VGG16. . . . .	12
2.4	Representação visual das camadas que constituem a rede VGG19. . . . .	12
2.5	Representação visual das camadas que constituem a rede InceptionV3. . . . .	13
2.6	Representação visual das camadas que constituem a rede ResNet50V2. . . . .	14
2.7	Representação do cálculo residual na ResNet50V2. . . . .	15
2.8	Ilustração do processo de aquisição de uma imagem de radiografia. . . . .	16
2.9	Representação da técnica validação cruzada com 5 divisões. . . . .	17
2.10	Imagem genérica de uma Matriz de Confusão. . . . .	18
3.1	Representação do dataset na primeira iteração. . . . .	20
3.2	Imagem de um paciente que contém aumento de átrio esquerdo, lateralidade esquerda. . . . .	21
3.3	Imagem de um paciente normal, lateralidade esquerda. . . . .	21
4.1	Matriz de confusão da vgg16 com learning rate = 0,001. . . . .	27
4.2	Média da loss da vgg16 com learning rate = 0,001. . . . .	27
4.3	Matriz de confusão da vgg16 com learning rate = 0,0001. . . . .	28
4.4	Média da loss da vgg16 com learning rate = 0,0001. . . . .	28
4.5	Matriz de confusão da vgg19 com learning rate = 0,001. . . . .	29
4.6	Média da loss da da vgg19 com learning rate = 0,001. . . . .	29
4.7	Matriz de confusão da vgg19 com learning rate = 0,0001. . . . .	31
4.8	Média da loss da da vgg19 com learning rate = 0,0001. . . . .	31
4.9	Matriz de confusão da inceptionV3 com learning rate = 0,001. . . . .	32
4.10	Média da loss da da inceptionV3 com learning rate = 0,001. . . . .	32
4.11	Matriz de confusão da inceptionV3 com learning rate = 0,0001. . . . .	33
4.12	Média da loss da da inceptionV3 com learning rate = 0,0001. . . . .	33
4.13	Matriz de confusão da resNet50V2 com learning rate = 0,001. . . . .	34
4.14	Média da loss da da resNet50V2 com learning rate = 0,001. . . . .	34
4.15	Matriz de confusão da resNet50V2 com learning rate = 0,0001. . . . .	35
4.16	Média da loss da da resNet50V2 com learning rate = 0,0001. . . . .	35

## LISTA DE TABELAS

3.1	Camadas customizadas adicionadas às redes criadas . . . . .	22
3.2	Camadas substituídas da VGG16 e VGG19 . . . . .	23
3.3	Camadas substituídas da InceptionV3 e Resnet50V2. . . . .	23
4.1	Resultados da média de quatro dias de execução da rede VGG16 com taxa de aprendizado 0,001. . . . .	25
4.2	Resultados da média de quatro dias de execução da rede VGG16 com taxa de aprendizado 0,0001 . . . . .	26
4.3	Resultados agregados do desempenho da VGG16 . . . . .	26
4.4	Resultados da média de quatro dias de execução da rede VGG19 com taxa de aprendizado 0,001. . . . .	26
4.5	Resultados da média de quatro dias de execução da rede VGG19 com taxa de aprendizado 0,0001 . . . . .	26
4.6	Resultados agregados do desempenho da VGG19 . . . . .	26
4.7	Resultados da média de quatro dias de execução da rede InceptionV3com taxa de aprendizado 0,001. . . . .	30
4.8	Resultados da média de quatro dias de execução da rede InceptionV3com taxa de aprendizado 0,0001 . . . . .	30
4.9	Resultados agregados do desempenho da InceptionV3. . . . .	30
4.10	Resultados da média de quatro dias de execução da rede ResNet50V2 com taxa de aprendizado 0,001 . . . . .	30
4.11	Resultados da média de quatro dias de execução da rede InceptionV3com taxa de aprendizado 0,0001 . . . . .	30
4.12	Resultados agregados do desempenho da ResNet50V2 . . . . .	30
4.13	Agrupamento dos melhores resultados por rede . . . . .	33

## LISTA DE ACRÔNIMOS

VGG16	Visual Geometry Group 16
VGG19	Visual Geometry Group 19
PACS	Picture Archiving and Communication Systems
Dicom	Digital Imaging and Communications in Medicine
ReLu	Rectified Linear Unit
RNA	Rede Neural Artificial
CNN	Convolutional Neural Network
LOO	Leave One Out

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>8</b>
1.1	PROPOSTA	8
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>9</b>
2.1	DIAGNÓSTICO ASSISTIDO POR COMPUTADOR	9
2.2	REDES NEURAIS	9
2.3	REDES NEURAIS CONVOLUCIONAIS	11
2.3.1	REDES NEURAIS PARA CLASSIFICAÇÃO	11
2.4	IMAGENS RADIOGRÁFICAS	14
2.4.1	DATA AUGMENTATION	16
2.4.2	VALIDAÇÃO CRUZADA	16
2.5	MÉTRICAS DE AVALIAÇÃO DOS MODELOS	17
2.6	TRABALHOS RELACIONADOS	18
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>20</b>
3.1	DATASET	20
3.2	DESIGN EXPERIMENTAL	20
3.2.1	Data augmentation	21
3.2.2	Técnicas aplicadas nos experimentos	22
3.3	EXPERIMENTO	22
3.3.1	Configurações iniciais das redes	23
3.3.2	Treino, Validação e Teste	23
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b>	<b>25</b>
4.1	VGG16	25
4.2	VGG19	25
4.3	INCEPTIONV3	27
4.4	RESNET50V2	28
4.5	ANÁLISE DOS RESULTADOS	29
<b>5</b>	<b>CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS</b>	<b>36</b>
	<b>REFERÊNCIAS</b>	<b>38</b>

# 1 INTRODUÇÃO

Conforme Atkins et al. (2009), cerca de 10% dos cães de estimação que passam por exames de rotina apresentam problemas cardíacos, sendo a doença Mixomatosa valvar mitral aquela que se demonstra presente em três a cada quatro pacientes que apresentem os problemas cardíacos. O desenvolvimento da doença tem como consequência o aumento do tamanho do átrio esquerdo dos animais, o que torna essa característica um forte fator para a identificação desses casos. A verificação da ocorrência desse sintoma é facilitada através de exames radiológicos, pois as imagens geradas nesses exames, a baixo custo, permitem aos profissionais identificarem as ocorrências da anomalia no tamanho do átrio. No entanto, o processo de análise e diagnóstico a partir da imagem radiográfica é um processo demorado e custoso para o profissional, o que torna interessante a perspectiva da adoção de uma ferramenta para auxiliá-lo a validar seus laudos sem custos de mão-de-obra adicionais.

## 1.1 PROPOSTA

Utilizando modelos de aprendizagem profunda para desempenhar atividades como segmentação de regiões de interesse ou mesmo classificar as imagens quanto à presença ou ausência do sintoma, algoritmos de aprendizado profundo podem prestar auxílio ao diagnóstico por meio de contribuições como o aumento de acurácia da diagnose e com ferramentas que auxiliem a interpretação dos resultados presentes nas imagens apresentarem maior consistência (Azevedo-Marques, 2001). Idealmente, espera-se que um sistema de auxílio ao diagnóstico eficiente possa também reduzir a quantidade de casos encaminhados aos profissionais, contribuindo positivamente para a eficiência do processo do diagnóstico. Sendo assim, considera-se interessante que modelos de aprendizado profundo tenham sua eficiência comparada em tarefas de classificação de imagens radiográficas, como normais ou doentes, para que seja possível caminhar em direção ao desenvolvimento de uma ferramenta que consiga prestar o auxílio ao diagnóstico eficientemente para os profissionais, reduzindo a complexidade do diagnóstico e agilizando o atendimento aos pacientes necessitados.

Dentre objetivos específicos deste trabalho, estão:

- Desenvolver ferramentas para a detecção da presença do aumento do átrio esquerdo nas imagens radiográficas e a aplicação para auxiliar o profissional como uma segunda opinião;
- Comparação do desempenho de modelos de aprendizado profundo para a classificação de imagens biomédicas, provindas de exames radiográficos;
- Avaliar o desempenho de modelos de aprendizado profundo em um ? com poucas amostras.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo contém informações acerca dos principais conceitos necessários para o desenvolvimento deste trabalho, abordando alguns tópicos sob a perspectiva estritamente computacional e alguns outros sob a perspectiva da intersecção da computação com a veterinária.

### 2.1 DIAGNÓSTICO ASSISTIDO POR COMPUTADOR

“Como os diagnósticos de radiologistas podem ser auxiliados pelos benefícios das imagens digitais?” foi o questionamento que deu origem ao *Computer Aided Diagnosis* (CAD) (Doi, 2007), ou diagnóstico assistido por computador. Como o próprio nome infere, sua finalidade é servir como uma “segunda opinião” ao profissional de radiologia quando houver necessidade da análise de uma imagem. Sendo um processo custoso, o profissional de saúde sofre desgaste ao efetivar a diagnose de quadros que necessitem a análise de imagens radiográficas, o que pode influenciar na correteude e na eficácia dos diagnósticos. Sendo assim, o uso de uma ferramenta computacional para confirmar hipóteses se mostra de grande valia. Conforme Azevedo-Marques (2001), o auxílio prestado se dá pela localização de regiões que contenham padrões radiológicos suspeitos em imagens, e a partir disso, pode-se partir para trabalhos de segmentação da região suspeita, para possibilitar análises direcionadas às suspeitas encontradas, ou classificação da imagem em questão sob a perspectiva de auxiliar a validação do laudo profissional.

Por outro lado, passar exames radiológicos por uma triagem, antes de fornecê-las ao profissional para a análise e diagnóstico, pode providenciar grande economia por evitar trabalho desnecessário e custoso do profissional, e é a partir da necessidade de certeza para classificar exames corretamente que surgem barreiras para a adoção da abordagem e trabalhos como este são fundamentados, porém ainda sob a perspectiva da não-automatização completa do processo de diagnóstico para que o profissional ainda esteja presente no processo.

### 2.2 REDES NEURAIS

Como pertencentes à área de aprendizado de máquina, as Redes Neurais, também chamadas de Redes Neurais Artificiais (RNA), são inspiradas no cérebro humano com o intuito de imitar o funcionamento das ligações nervosas que nele se estabelecerem<sup>1</sup>. Sua topologia, ou arquitetura, é composta pelas denominadas camadas, cada uma contendo um conjunto de neurônios além de especificações de entrada e saída para que seja possível concatená-las umas às outras. Sua configuração mais clássica consiste em uma camada de entrada, uma sequência de camadas escondidas e uma camada de saída, conforme a Figura 2.1. Tradicionalmente, os dados percorrem somente uma direção entre as camadas que compõe a rede.

Conforme em e Silva e Schmidt (2016) e ilustrado na Figura 2.2, um neurônio artificial é composto pelos seguintes componentes: sinais de entrada  $X_m$ ; pesos (ou pesos sinápticos)  $W_{kp}$ ; bias  $b_k$ ; somatório dos pesos representado por  $\Sigma$ ; combinação linear dos sinais de entrada  $u_k$ ; função de ativação  $\varphi(\cdot)$  e a saída  $y_k$ . O processo da repetição da atribuição de pesos a um neurônio, visando que a saída  $y_k$  obtenha um valor esperado, caracteriza o processo de aprendizado de um neurônio para uma determinada entrada, de forma que esse valor resultante nos pesos gerem o valor de saída persistentemente para a entrada utilizada no treino. Utilizar

<sup>1</sup><https://www.ibm.com/topics/neural-networks>

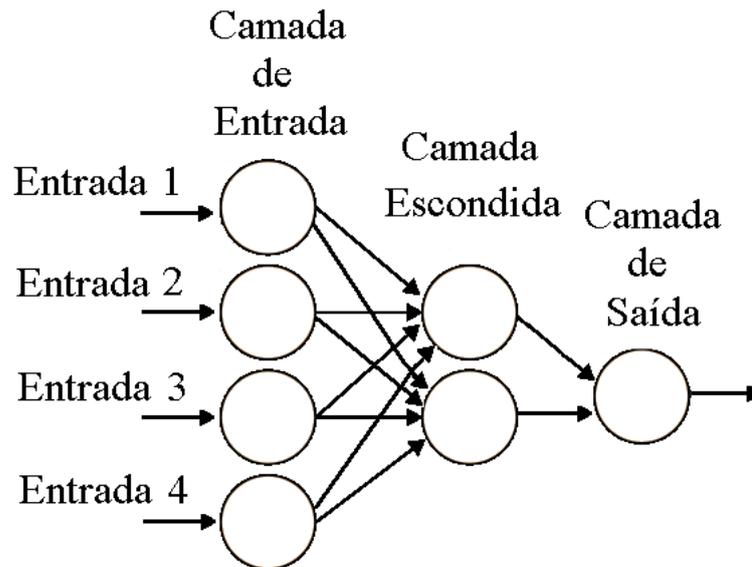


Figura 2.1: Representação da topologia de uma rede neural artificial.

Fonte: Os autores

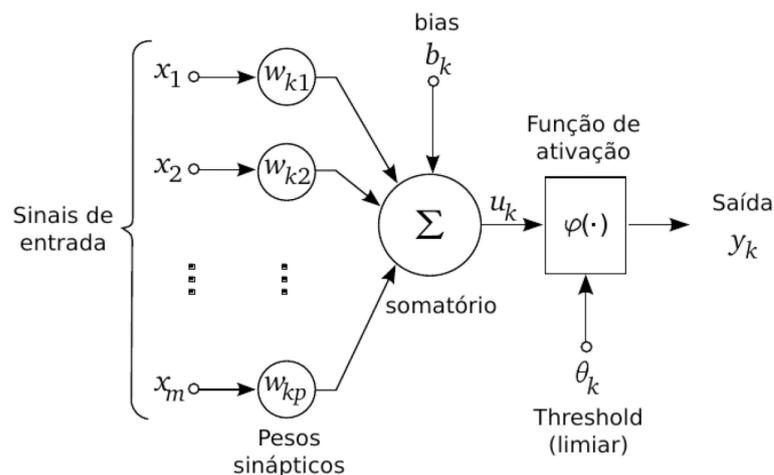


Figura 2.2: Representação do neurônio de uma rede neural artificial.

Fonte: (e Silva e Schmidt, 2016)

uma quantidade de neurônios que sejam interligados entre si, com a saída de um alimentando a entrada de outra, é o que forma uma rede neural artificial, como o demonstrado na Figura 2.1.

Sendo assim, em tarefas que necessitem valores de saídas para diferentes valores de entrada, o processo de treino deverá ser aplicado para cada uma das entradas disponíveis visando a adequação dos pesos. A generalização, ou seja, a atribuição de um valor de saída conhecidamente correto para uma amostra que não pertença ao conjunto de treino, é possibilitada pelo uso de grandes volumes de dados para que seja possível adquirir pesos representativos para a caracterização das amostras treinadas. A eficiência do processo é, então, diretamente influenciada pela representatividade e diversidade dos dados fornecidos para o treino.

Os dados de treinamento de uma Rede Neural devem ser a representação adequada das instâncias referentes ao problema em que ela será aplicada, tal como em problemas que envolvam textos, os dados consistirem em valores que representem as palavras que o compõe, enquanto em problemas que envolvam imagens, os dados são os valores que constituem os pixels delas.

## 2.3 REDES NEURAS CONVOLUCIONAIS

Redes Neurais Convolucionais são uma subcategoria das Redes Neurais, geralmente voltada para problemas de reconhecimento de padrões, visão computacional e reconhecimento de imagens<sup>2</sup>. Conforme Lin et al. (2013), são a alternância entre camadas convolucionais e de agregação, podendo-se destacar essa como sua principal distinção das Redes Neurais previamente citadas. Uma convolução, a operação que ocorre nas camadas convolucionais, é de um produto entre duas matrizes, que consistem nos dados de entrada da camada sendo multiplicados pelo filtro dela cujos valores também podem ser alterados no processo de treino da rede. Os resultados fornecidos por uma camada convolucional podem ser entendidos como de mapas de características, que contém as informações condensadas das características extraídas da matriz recebida como entrada para a camada convolucional. Conforme os mapas de características passem por mais camadas convolucionais as informações armazenadas nos mapas de características passarão de formas genéricas para serem representações do objeto discriminatório em si, tal como bordas presentes nas imagens que compõe as amostras fornecidas à rede.

As camadas convolucionais são utilizadas em conjunto com funções de ativação, camadas de agregação e camadas completamente conectadas. Funções de ativação servem para introduzir não-linearidade nos mapas de características; as camadas de agregação para que as características com mais poder de discriminação sejam priorizadas para passarem adiante no processamento da rede; e as camadas completamente conectadas servem para fazer a classificação, propriamente dita, da amostra processada<sup>3</sup>.

### 2.3.1 REDES NEURAS PARA CLASSIFICAÇÃO

A discriminação da amostra em uma das possíveis classes para o problema, em uma Rede Neural Convolucional, conforme previamente descrito, ocorre nas camadas completamente conectadas. Elas calculam níveis de certeza de pertença da amostra para as possíveis classes cabíveis ao problema, de forma que a classificação a ser atribuída seja aquela com maior certeza.

Neste trabalho foram selecionados quatro modelos para a criação de Redes Neurais customizadas a fim de classificar as imagens entre aquelas que apresentam o paciente com aumento de átrio e aquelas que apresentam pacientes sem o sintoma. Os modelos escolhidos foram VGG16, VGG19, InceptionV3 e ResNet50v2.

#### 2.3.1.1 VGG16

Desenvolvida pelo Grupo de Geometria Visual da Universidade de Oxford, a VGG16 obteve desempenho marcante em sua apresentação na ILSVRC (ImageNet Large-Scale Visual Recognition Challenge), por considerar aspectos que anteriormente não recebiam a devida atenção pelas redes convolucionais para a extração de características dos dados: a profundidade da rede e o tamanho do filtro das convoluções. Para isso, com os demais parâmetros fixados, a profundidade da rede foi aumentada pela adição de camadas convolucionais, em conjunto com o uso de menores tamanhos de filtros (3x3) para a realização das convoluções (Simonyan e Zisserman, 2014) e pela fixação das dimensões de entradas aceitas. O tamanho reduzido para os filtros das convoluções foi utilizado em todas as ocorrências de camadas convolucionais no modelo, assim como o *stride* e o *padding* das operações convolucionais foram fixados em 1. Para as ocorrências da operação de agregação máxima, após os blocos de camadas convolucionais, são utilizados *strides* de valor 2 e janelas de pixels de dimensões 2x2. Para as ativações, a função

<sup>2</sup><https://www.ibm.com/topics/neural-networks>

<sup>3</sup><https://www.ibm.com/topics/convolutional-neural-networks>

sigmoide foi a escolhida. Após sequência de convoluções, *poolings* e ativações, conforme a Figura 2.3, obtém-se um vetor composto pelas características extraídas no processo de aprendizado da rede. Então, para a classificação, são utilizadas as camadas completamente conectadas da rede, que terão a função de, dadas as características extraídas até então, calcular a probabilidade de pertinência a cada uma das possíveis classes do problema abordado para a amostra fornecida. A atribuição a uma determinada classe ocorre para a classe que tenha maior probabilidade calculada dentre aquelas possíveis.

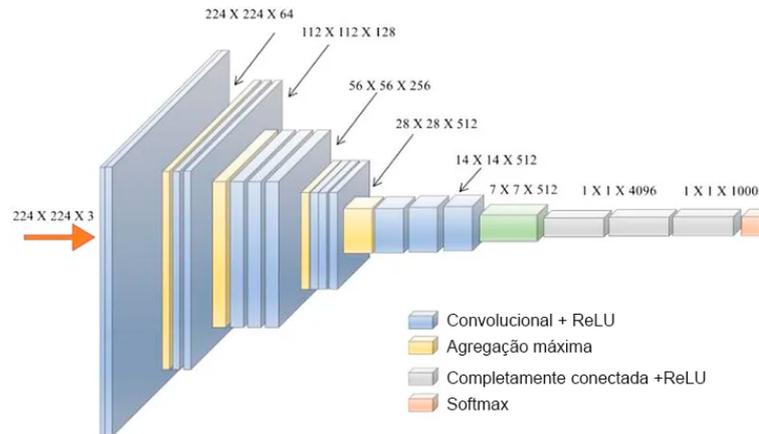


Figura 2.3: Representação visual das camadas que constituem a rede VGG16.  
Fonte: Os autores

### 2.3.1.2 VGG19

De forma similar e de mesma autoria que a VGG16, a proposição da rede VGG19 seguiu o mesmo princípio da VGG16 referente ao que diz respeito à profundidade da arquitetura e ao tamanho das convoluções. Muito similares, o que distingue a VGG19 da VGG16 é a adição de mais três camadas convolucionais, uma em cada um dos três últimos blocos de camadas convolucionais (Figura 2.4), que anteriormente eram compostos por apenas 3 camadas convolucionais antes da agregação máxima.

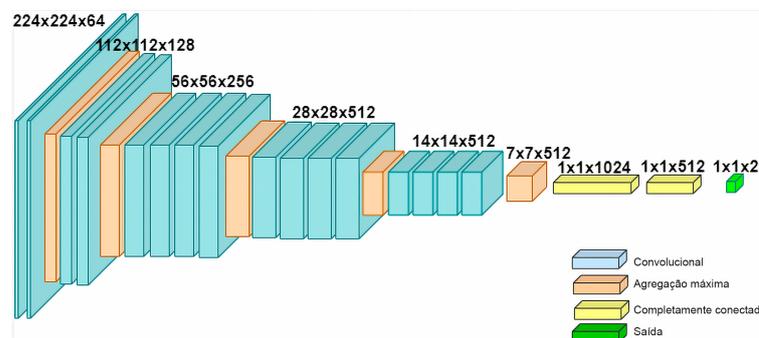


Figura 2.4: Representação visual das camadas que constituem a rede VGG19.  
Fonte: Os autores

### 2.3.1.3 InceptionV3

Tanto a rede VGG16 quanto a rede VGG19 têm como uma grande marca a simplicidade de suas arquiteturas, que consistem em fluxos lineares com um começo e um fim bem-definidos.

No entanto, tal simplicidade de compreensão e uso deve ser compensada de alguma forma, e neste caso, tal forma é a quantidade de parâmetros treináveis existentes. A maior concentração deles ocorre nas camadas densas próximas ao término da rede, especificamente por serem responsáveis pela atribuição de uma classe às características recebidas como entrada. Em Szegedy et al. (2016), é destacado por integrantes da Google que, além do número de parâmetros, a simplicidade de compreensão da arquitetura torna o custo para a avaliação da rede mais caro também. Com base nisso, como uma proposta originada desse trabalho, a InceptionV3 tem como algumas de suas características possuir uma arquitetura menos linear, reduzir significativamente o número de parâmetros treináveis da rede e a facilidade para sua aplicação em problemas de *big-data*, devido às tentativas de manter o custo da computação da rede baixa enquanto melhorando a qualidade dos resultados.

Conforme a figura Figura 2.5, a InceptionV3, baseada nos modelos da GoogleLeNet, é constituída por camadas convolucionais, de *pooling* médio, *pooling* máximo, concatenantes, *dropout*, densas e de ativação Softmax. As camadas concatenantes são novidade relativamente aos modelos previamente citados, e sua função é ponderar os resultados das convoluções executadas em paralelo em diferentes etapas da rede, sendo esse paralelismo uma outra inovação da rede. Para a realização das convoluções, o conceito de fatorização foi intriduzido, que consiste na decomposição de uma operação convolutiva, de forma que o resultado de mais convoluções, sequenciais entre si e com filtros menores, seja similar ao de uma convolução efetuada com um filtro maior. Por exemplo, a substituição de uma camada convolucional com um filtro de dimensões 7x7 por duas camadas convolucionais ambas com filtros de dimensões 3x3. Szegedy et al. (2016) também obtiveram medições que indicam que as convoluções fatoradas acabam sendo mais baratas do que uma convolução não fatorada, além da constatação de que a decomposição de, por exemplo, uma convolução com filtro 3x3 em duas novas, de filtros 3x1 e 1x3, não seria vantajosa. Por fim, também é válido destacar que foi adotado o conceito de resultados parciais, na tentativa de auxiliar na convergência dos resultados gerados pela rede.

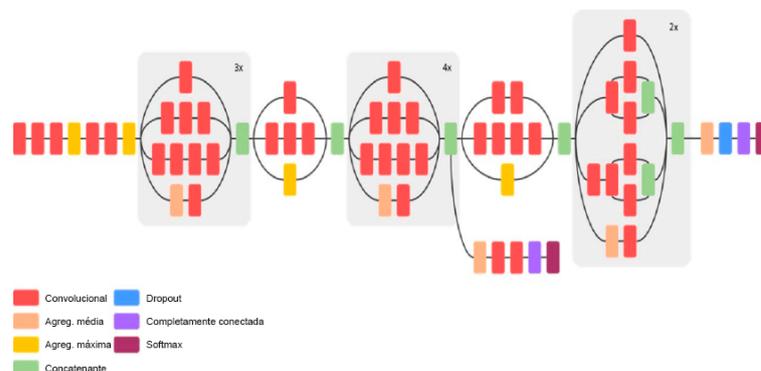


Figura 2.5: Representação visual das camadas que constituem a rede InceptionV3.

Fonte: Os autores

#### 2.3.1.4 ResNet50V2

Conforme anteriormente mencionado e em conjunto com as Figuras 2.3, 2.4 e 2.5, percebe-se que a adição de mais camadas para aumentar a profundidade é uma tendência para a tentativa de obtenção de redes com melhor desempenho na classificação de imagens. No entanto, (He et al., 2016) passa a considerar que, conforme se tornem mais profundas, as redes tendem a ter sua convergência degradada (devido à saturação da acurácia). A partir disso, a adição de novas camadas apenas acarretaria maiores erros de treino e surtiriam efeito contrário ao esperado

até então. Na tentativa de contornar o efeito degradante, He et al. (2016) então a introdução do uso de resíduos na arquitetura das redes, ou seja, o reaproveitamento de resultados anteriores em pontos mais profundos da rede na tentativa de preservar características adquiridas no processo de extração enquanto o processo de refinamento delas ainda é continuado. O uso dos resíduos ocorre por meio das nomeadas 'conecções atalho', que consistem em ligações entre uma camada a uma próxima com pelo menos dois níveis de profundidade a mais e efetua, originalmente, a soma do resíduo com o resultado obtido no processamento que está sendo 'pulado' pela conexão atalho.

A ResNet50V2 consiste, então, em uma versão das redes de aprendizado residual constituída por 50 camadas e uma mudança na forma com que o resíduo é agregado aos resultados posteriores da rede em relação à ResNet50. Quanto às suas camadas, a rede é constituída por 48 convolucionais, uma de agregação média, uma de agregação máxima, além de um bloco de contendo camadas densas completamente conectadas, conforme a Figura 2.6. Já para a agregação do resíduo aos resultados posteriores, a mudança foi no cálculo do resíduo carregado na rede, visando que ele se torne mais representativo (ilustrado na Figura 2.7).

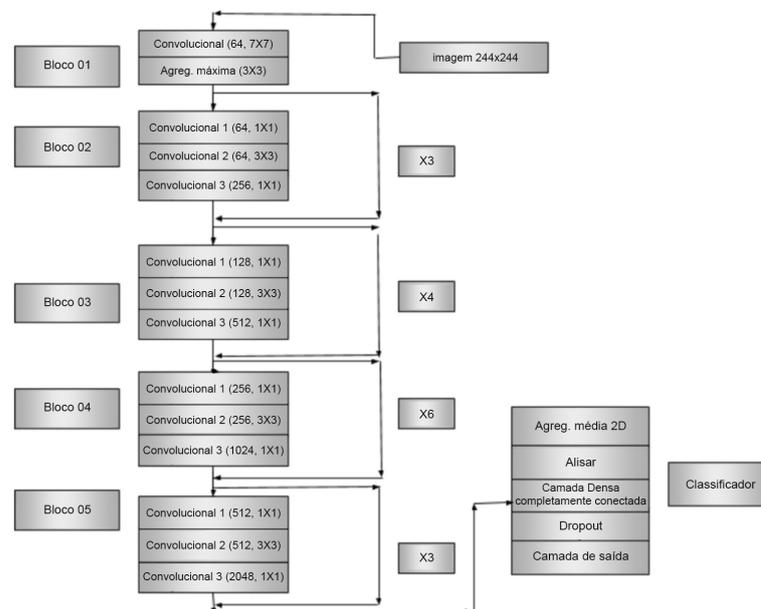


Figura 2.6: Representação visual das camadas que constituem a rede ResNet50V2.

Fonte: Os autores

## 2.4 IMAGENS RADIOGRÁFICAS

Uma imagem radiográfica é uma imagem provinda de um exame de Raio-x, que é utilizado na medicina, tanto humana quanto veterinária, como uma forma de permitir que hipóteses de ocorrências de lesões ou enfermidades nos pacientes sejam validadas ou descartadas de forma não-evasiva e de baixo custo.

A formação das imagens provindas desses exames radiológicos ocorre por meio da interação da radiação emitida pelo aparelho de Raios-x com a matéria, correspondente ao corpo do paciente que esteja fazendo o exame. Tal interação possui dependência conforme fatores como o foco dos feixes de radiação emitidos, distância até o paciente e especificações técnicas do aparelho utilizado.

Os feixes de radiação, dimensionados pelo colimador, são bombardeados no paciente, atingindo as estruturas alvejadas que são separadas entre duas classificações: radiopacas ou

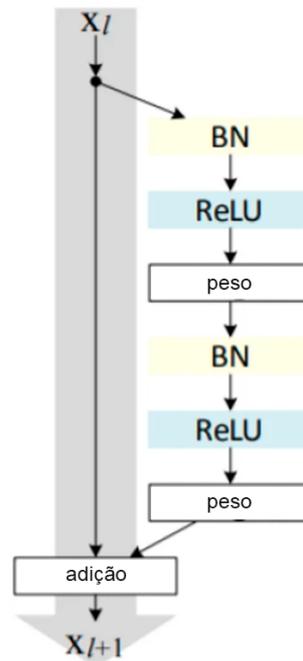


Figura 2.7: Representação do cálculo residual na ResNet50V2.

Fonte: Os autores

radiolúcidas. Estruturas radiolúcidas, como a pele, absorvem em baixa quantidade a radiação, permitindo que ela consistentemente atravesse os meios que atinge. Estruturas radiopacas são aquelas que são mais densas, que têm maior interação com a radiação incidente, como principalmente os ossos. Nelas, espera-se que o efeito fotoelétrico ocorra, pois é o responsável pela maior parte da formação da imagem.

Os receptores posicionados atrás do paciente irão direcionar a radiação incidente ao filme onde será formada a imagem, em tempo de exame. A Figura 2.8 mostra todos os componentes na aquisição de uma imagem de Raio-X. A imagem resultante consiste em regiões cuja intensidade varia conforme a quantidade de radiação incidente o aparelho recebe. Regiões mais escuras denotam maior incidência de radiação no receptor, enquanto as mais claras denotam uma menor incidência. Essas também são estruturas responsáveis pela maior influência na ocorrência de ruídos nas imagens resultantes, pois o processo físico da absorção da radiação envolve a liberação de energia em ângulos diferentes ao de incidência. A adoção de colimadores, dentre outras técnicas e dispositivos, também no receptor auxiliam a reduzir a geração de ruídos, pois filtram os feixes incidentes conforme seus ângulos de incidência, de forma que a radiação causadora de ruídos seja impedida de interferir na formação da imagem. Estruturas como o coração e o pulmão encontram-se em um tom acinzentado, resultante da absorção parcial da radiação, e a ocorrência de tumores e deformações mostram-se perceptíveis por possuírem densidades e superfícies de contato diferentes das regiões que não as contém, influenciando na radiação que chega ao receptor.

A imagem resultante do processo de aquisição é armazenada conforme o padrão DICOM. Nele, além de informações da imagem, estão contidos metadados referentes ao paciente e ao exame de aquisição (como informações do aparelho utilizado, por exemplo). O uso desse formato é importante para viabilizar e facilitar a transferência de dados por meio do PACS entre aparelhos cujas versões não precisem ser padronizadas.

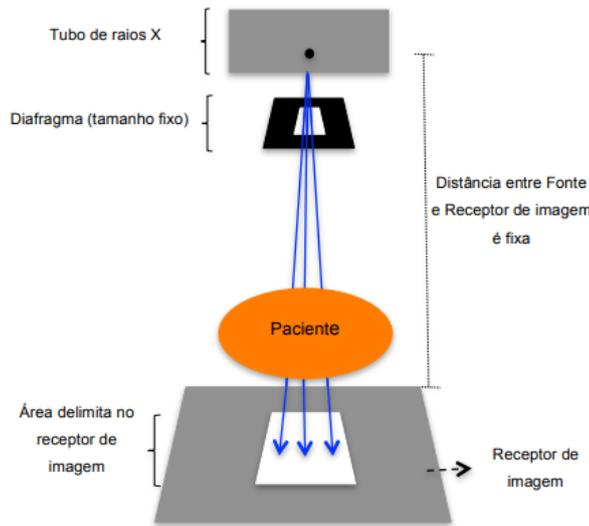


Figura 2.8: Ilustração do processo de aquisição de uma imagem de radiografia.

Fonte: Princípios Físicos em Radiologia

#### 2.4.1 DATA AUGMENTATION

*Data augmentation* consiste em uma técnica para, artificialmente, aumentar o número de dados do conjunto de treino de uma rede neural, por meio de transformações sobre os dados originais (Awan, 2022). No escopo deste trabalho, consiste na aplicação de operações matemáticas para modificar as imagens radiográficas, na tentativa de compensar a baixa quantidade de imagens disponíveis. Conforme Perez e Wang (2017), transformações comumente aplicadas no processo de augmentation são geométricas ou transformações de cor. Quanto às geométricas são elas por exemplo recorte, espelhamento e translação; para as de cor, envolvem mudanças na paleta de cores da imagem. Neste trabalho por se tratarem de imagens biomédicas não foram aplicadas transformações que modifiquem drasticamente a imagem, pois o espelhamento (vertical ou horizontal), por exemplo, de uma imagem em um determinado plano de aquisição pode descaracterizá-la completamente uma vez que a 'orientação' seja uma possível característica de peso para a classificação.

#### 2.4.2 VALIDAÇÃO CRUZADA

A técnica da validação cruzada se trata de uma metodologia para o treino de redes neurais, especialmente utilizada em casos cujos *datasets* disponíveis para o treino não sejam volumosos ou representativos o suficiente. Visualmente representado na Figura 2.9, o método consiste em, dado um valor  $k$  (número de pastas escolhido), o conjunto de dados será dividido entre  $k$  grupos, que será mantida até o final da aplicação do método.

Efetivada a divisão dos dados, uma iteração da validação cruzada irá selecionar uma das pastas para compor o conjunto de dados de teste, deixando todas as  $k - 1$  restantes como o conjunto de treino para a rede, dentre os que poderão ser selecionados um subconjunto para compor os dados para a validação do treino. Terminada a iteração, uma nova pasta será selecionada para ser o conjunto de teste, diferente de todas as que já tenham sido utilizadas até então, de forma que o conjunto de treino seja ainda composto por  $k - 1$  pastas que não estejam sendo utilizadas para o teste. Sendo assim, é garantido que todos os dados serão utilizados para o conjunto de testes uma única vez no processo de treino, evitando que os pesos selecionados ao final do processo sejam tendenciosos a uma amostra em específico do conjunto.

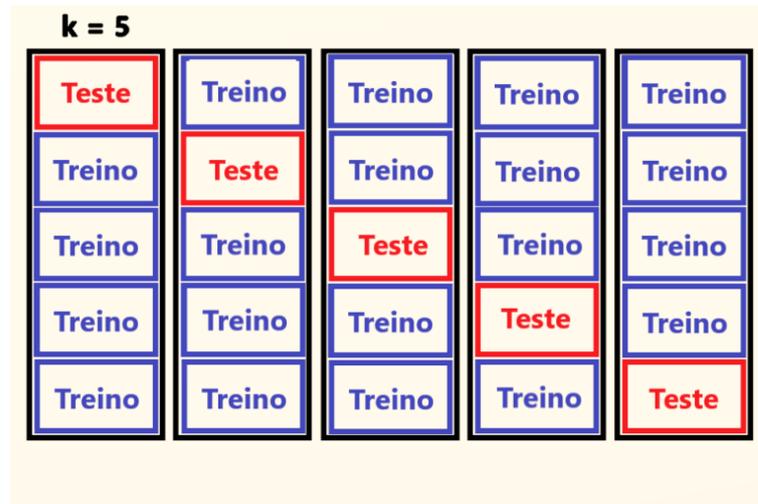


Figura 2.9: Representação da técnica validação cruzada com 5 divisões.

Fonte: (Tech, 2022)

Um caso particular do método da validação cruzada se trata de quando o número de pastas é exatamente igual ao número de amostras, fazendo com que uma única amostra seja o conjunto de teste de cada iteração. Nesses casos, a metodologia é referenciada como *Leave One Patient Out*, quando aplicada à contextos que lidem com dados relacionados a saúde, ou simplesmente *Leave One Out* (LOO). Os casos em que o LOO é aplicado costumam ser aqueles em que o dataset seja demasiadamente reduzido, consistindo em uma tentativa de adequar o processo de treino da rede a casos extremos.

## 2.5 MÉTRICAS DE AVALIAÇÃO DOS MODELOS

O processo de organização e análise dos resultados de experimentos de classificação entre múltiplas classes envolve a criação de matrizes de confusão, que consistem na relação entre a quantidade de ocorrências de resultados atribuídos a cada uma das possíveis classes previstas no escopo de um problema. Como resultado, a diagonal principal da matriz torna-se composta pelos verdadeiros positivos e as outras entradas correspondem à quantidade de ocorrências errôneas entre as classes. Em contextos que sejam utilizadas apenas duas classes, a matriz será constituída por quatro entradas: quantidade de imagens normais classificadas corretamente, quantidade de imagens normais classificadas como contendo aumento de átrio, quantidade de imagens que contém aumento de átrio classificadas corretamente e quantidade de imagens com aumento de átrio classificadas como normais.

Para avaliar a qualidade dos resultados produzidos pelo processo de treino, teste e validação das redes foram utilizadas as seguintes métricas: acurácia, sensibilidade, especificidade, precisão e F1-Score. Para a caracterização de tais métricas, são considerados os possíveis resultados do processo de classificação: verdadeiros positivos (VP), como predições corretas de que uma imagem contém aumento de átrio esquerdo; verdadeiros negativo (VN), como predições corretas de que uma imagem é de um paciente normal; falso negativo (FN), como a classificação da imagem como normal enquanto ela contenha um paciente com aumento de átrio; e por fim falso positivo (FP), como a classificação da imagem como contendo aumento de átrio enquanto ela seja normal.

Com isso, pode-se definir as métricas previamente citadas conforme as seguintes fórmulas:

		Valor Predito	
		Sim	Não
Real	Sim	Verdadeiro Positivo (TP)	Falso Negativo (FN)
	Não	Falso Positivo (FP)	Verdadeiro Negativo (TN)

Figura 2.10: Imagem genérica de uma Matriz de Confusão.

Fonte: (Nogare, 2020)

- Acurácia =  $\frac{VP+VN}{VP+VN+FP+FN}$
- Sensibilidade =  $\frac{VP}{VP+FN}$
- Especificidade =  $\frac{VN}{FP+VN}$
- Precisão =  $\frac{VP}{VP+FP}$
- F1-Score =  $\frac{2 \times \text{Acurcia} \times \text{Sensibilidade}}{\text{Acurcia} + \text{Sensibilidade}}$

A combinação das métricas são consideradas para que conclusões menos tendenciosas possam ser tiradas. Por se tratar de um contexto de auxílio a diagnóstico, sensibilidade e especificidade são fortemente consideradas para a realização de inferências sobre o resultado, uma vez que resultados errôneos podem induzir a decisões que levem o paciente a tratamentos desnecessários ou que nenhum auxílio seja prestado em casos que sejam necessários. Tais erros não são comparáveis quanto à sua gravidade, uma vez que ambos possuam graves consequências ao paciente.

## 2.6 TRABALHOS RELACIONADOS

Para o desenvolvimento deste trabalho, alguns artigos da área foram estudados e serão mencionados brevemente nesta sessão. Todos lidam com problemas relacionados ao treino de modelos de aprendizado profundo com imagens radiográficas para a classificação entre as possibilidades cabíveis.

Tomando como base trabalhos cujos objetos de estudo são imagens radiográficas de caninos para o treino de redes neurais, tem-se que Li et al. (2020), utilizando um *dataset* composto por 711 imagens para treino e 81 para teste, foi aplicada a técnica da validação cruzada para experimentar o desempenho da rede VGG16 para a classificação das amostras. Os testes giraram em torno da avaliação do desempenho do modelo em dois casos, sendo o primeiro em que o treino da rede prioriza especificidade e sensibilidade igualmente enquanto o segundo prioriza melhorias na sensibilidade em relação à especificidade.

Já Banzato et al. (2021a) propõe uma comparação entre as redes ResNet50 e DenseNet121. Também utilizando o formato jpeg para as imagens, o processo de treino e validação é executado

entre dois *datasets* distintos que somados totalizam aproximadamente 3800 imagens. Limitações deste trabalho envolvem a composição do *dataset* utilizado por imagens providas de apenas uma instituição, o que limita a qualidade dos resultados, conforme os autores. Com autoria similar, Banzato et al. (2021b) utiliza radiografias apenas de felinos para avaliar o desempenho das redes ResNet50 e InceptionV3, treinando-as em 1062 imagens para um problema de classificação multiclasse.

Das et al. (2021), por outro lado, utilizando imagens de radiografias em humanos, estuda o agrupamento de três redes para que, utilizando a combinação por média ponderada, seja experimentada a qualidade dos resultados das três redes em conjunto. São elas DenseNet201, ResNet50v2 e Inceptionv3. O *dataset* utilizado totaliza 1006 amostras.

Por fim, Khan e Aslam (2020) utiliza um *dataset* de 1683 imagens radiográficas para a experimentação das redes DenseNet121, ResNet50, VGG16 e VGG19. O estudo consiste na comparação das redes citadas entre elas mesmas e, posteriormente, com dados encontrados pelos autores em um levantamento bibliográfico prévio.

### 3 MATERIAIS E MÉTODOS

Os materiais utilizados para a realização do trabalho foram as imagens radiográficas disponibilizadas para o estudo (3.1) e o ambiente de execução do Google Colab. Os métodos utilizados foram referentes a adequação do processo de treino de uma rede neural a baixos volumes de dados.

#### 3.1 DATASET

A aquisição das imagens ocorreu nas dependências do Laboratório de Imagens do Hospital Veterinário da Universidade Federal do Paraná, sendo composto de radiografias de caninos (cães) de classes normal (N), com aumento de átrio esquerdo (AAE) e com aumento de átrio esquerdo e edema pulmonar (AAEEP). Para os experimentos deste trabalho, foram apenas utilizadas as imagens das classes normal (N) e com aumento de átrio esquerdo (AAE). Para isso, todas as imagens passaram por processos de remoção de marcações, anonimização de dados sensíveis sob as recomendações do Comitê de Ética e Pesquisa e conversão do formato Dicom para Png. Mesmo contendo imagens sob a perspectiva de ambos os planos sagitais, foram utilizadas apenas aquelas com perspectiva do lado esquerdo do paciente, visando consistência no processo de aprendizado. O *dataset* utilizado possui 60 imagens sendo 30 de cães sem aumento do atrio e 30 com aumento. A Figura 11 contém um exemplo de uma das imagens utilizadas no experimento para um paciente doente, enquanto a Figura 12 contém o de um paciente normal.



Figura 3.1: Representação do dataset na primeira iteração.

Fonte: Os autores

#### 3.2 DESIGN EXPERIMENTAL

O experimento consiste na instanciação e modificação dos quatro modelos de redes neurais descritos previamente. Elas serão utilizadas conforme técnicas para treino, teste e validação de melhor desempenho em conjuntos de dados compostos por poucas amostras validação, para a classificação entre as duas possíveis classes (Normal e Aumento de Átrio Esquerdo) das amostras que compõe o *Dataset*. Após o processo, repetido quatro vezes visando a validação da consistência do experimento formulado, os resultados coletados foram agrupados e avaliados conforme a performance dos modelos utilizados. Tais validações foram executadas no período de quatro dias espaçados devido às limitações de recursos fornecidos pela plataforma do Google Colab. As configurações do ambiente são referentes ao plano “Pro+”, pois os outros não foram suficientes para a execução do experimento, consistindo em uma GPU Tesla T4 com 16GB de VRAM, e com versão do CUDA atualizada em 11.6.



Figura 3.2: Imagem de um paciente que contém aumento de átrio esquerdo, lateralidade esquerda.  
Fonte: Os autores

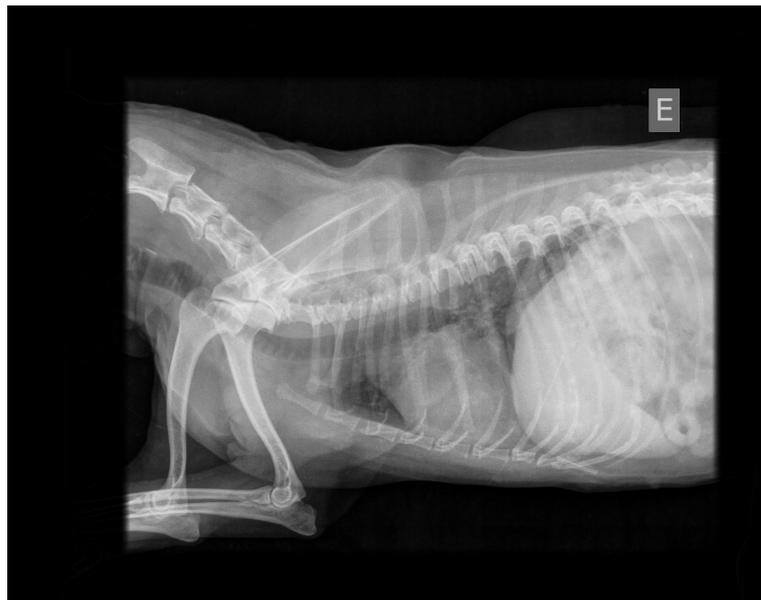


Figura 3.3: Imagem de um paciente normal, lateralidade esquerda.  
Fonte: Os autores

### 3.2.1 Data augmentation

Para o data augmentation, visando adquirir mais amostras para o processo de treino e aumentar a variabilidade desse conjunto, as quatro operações selecionadas para que o processo de *data augmentation* seja composto são as seguintes:

- Rotação entre  $15^\circ$  e  $-15^\circ$ ;
- Translação em entre 50 e -50 pixels;
- Redução no brilho em 10% do original;
- Mudanças no contraste em faixas aleatórias, variando de 20% a 100% do original.

Customizada
Agregação Máxima 2D Densa(512, ativação ReLu) Dropout(taxa 0, 5) Densa(2, ativação Sigmóide)

Tabela 3.1: Camadas customizadas adicionadas às redes criadas

Com esses processos, é possível prover leves modificações nas amostras sem prejuízo quanto as suas classes para que componham um *dataset* de treino com maior variabilidade em relação ao original.

### 3.2.2 Técnicas aplicadas nos experimentos

A fim de adquirir métricas para que os modelos previamente citados tenham seu desempenho avaliado e comparado no experimento, ocorreram divisões do *dataset* inicial em três porções. A primeira é utilizada para o treino dos modelos, a segunda para a validação do treino enquanto a terceira é reservada para testar os resultados do treino, possibilitando assim que métricas como acurácia, sensibilidade e f1-score sejam geradas para a comparação do desempenho entre os diferentes modelos avaliados.

Para o treino das redes, foram variadas as taxa de aprendizado em diferentes iterações de treino e validação parciais, que serão descritas mais detalhadamente posteriormente. A variação ocorreu por utilizar a taxa de aprendizado com os valores 0,001 e 0,0001 entre diferentes iterações. Para obter consistência nos resultados, a modificação dos hiperparâmetros somente é feita após a validação de uma das iterações do experimento, de forma que as métricas geradas em uma iteração qualquer sejam provindas de somente uma combinação dos hiperparâmetros. Além disso, alguns métodos para o treino e validação do desempenho foram utilizados, e serão descritos a seguir.

## 3.3 EXPERIMENTO

O código do experimento foi escrito na linguagem de programação Python (versão 3.8.10), no ambiente de desenvolvimento Google Colab - uma plataforma que aloca recursos remotamente para que códigos sejam executados via processamento em nuvem. A escolha foi feita para que ambos os autores pudessem acessar o código e os resultados simultaneamente e remotamente. Para a definição das redes a biblioteca Keras (versão 2.11.0) foi utilizada, pois é uma API que visa simplificar o uso dos recursos da biblioteca Tensorflow.

Definidas as redes, as últimas camadas, comumente responsáveis pela classificação dos dados, não foram definidas como as padrões oferecidas pela biblioteca. Foram adicionadas às redes uma camada de agregação máxima para a redução da dimensionalidade, uma densa com ativação ReLu, *Dropout* com a taxa em 0, 5 e mais uma densa, responsável pela classificação propriamente dita conforme o problema de classificação abordado no escopo do experimento. As substituições são ilustradas conforme as Tabelas 1, 2 e 3. A Tabela 1 contém as camadas adicionadas a cada uma das redes, enquanto as tabelas 2 e 3 contém as camadas que deixaram de ser utilizadas de cada uma das redes.

Os dados utilizados para o processo de treino, teste e validação delas foram do *dataset* balanceado composto por 60 imagens distribuídas igualmente entre cada uma das possíveis classes do problema, com 50% = 30 imagens da classe Normal e 50% = 30 imagens da classe

VGG16	VGG19
Alisamento	Alisamento
Densa(4096, ativação ReLu)	Densa(4096)
Densa(4096, ativação ReLu)	Densa(4096)
Densa(1000, ativação SoftMax)	Densa(1000, ativação SoftMax)

Tabela 3.2: Camadas substituídas da VGG16 e VGG19

InceptionV3	ResNet50V2
Agregação Global Média 2D	Agregação Global Média 2D
Densa(1000, ativação SoftMax)	Densa(1000, ativação SoftMax)

Tabela 3.3: Camadas substituídas da InceptionV3 e Resnet50V2

Átrio Aumentado. Sendo assim, para cada uma das redes o processo descrito a seguir foi aplicado, utilizando todas as 60 imagens disponíveis.

### 3.3.1 Configurações iniciais das redes

Ao princípio da execução do experimento com uma das quatro redes pré-determinadas elas são instanciadas e têm seus pesos atribuídos como aqueles provindos do pré-treino na base de dados *ImageNet*, fornecidos pela própria biblioteca Keras.

### 3.3.2 Treino, Validação e Teste

Variando a taxa de aprendizado do otimizador do modelo para os valores 0,001 e 0,0001, foram realizadas um total de  $60 \times 2 = 120$  iterações para cada um dos modelos previamente descritos, conforme a seleção de um dentre os possíveis valores da taxa de aprendizado e pela execução do método validação cruzada. O tamanho das pastas escolhidas para a aplicação do método foi  $k = 60$ , ou seja, com o tamanho de pastas igual ao número de amostras, somando então 60 iterações e garantindo que todos os dados disponíveis sejam percorridos para o uso no conjunto de teste. Para a compilação dos modelos, o otimizador que terá a learning rate variada é o "Adam" e a função de *loss* utilizada é a "*sparse categorical crossentropy*" conforme (Banzato et al., 2021b). A métrica avaliada para essas funções foi a acurácia. A priorização dessa métrica para a experimentação é devida à expectativa de que assim sejam aumentadas a especificidade e a sensibilidade em mesmas proporções para a classificação das amostras. A priorização de uma dessas métricas para os resultados do modelo varia entre a necessidade que o veterinário necessite para o problema abordado. Para este trabalho, nenhum parâmetro foi definido nesse sentido, por isso optou-se por comparar os resultados de uma maneira mais genérica.

Sendo assim, para uma iteração  $i$ , é separada uma amostra das 60 disponíveis para servir como teste ao final da iteração. Dentre as 59 restantes, são separadas outras 6 imagens do *dataset* para servir como o conjunto de validação. Dadas as  $60 - 1 - 6 = 53$  imagens restantes, é aplicado o processo de *data augmentation* conforme as 4 operações previamente descritas, para que o conjunto de imagens destinadas ao treino da rede seja composto por  $53 \times 5 = 265$  amostras.

Após o aumento dos dados de treino, utilizando os métodos previstos e oferecidos pelo Keras, o treino, teste e validação da rede são efetuados. É especificado para o processo de treino que sejam utilizadas 20 épocas. Para o teste, é utilizado o parâmetro *callbacks*, e nele é passada uma instância da função "*ModelCheckpoint*". Esse parâmetro é responsável pela geração de *checkpoints* no processo de teste, para que os melhores pesos sejam armazenados em arquivos de

extensão `.h5`; cujos caminhos são passados como parâmetros durante a instanciação da função, para que sejam utilizados em iterações futuras como os valores dos pesos iniciais das redes. Cada rede possui seu próprio conjunto de pesos, de forma que o treino de uma delas não interfira na outra.

### 3.3.2.1 *Geração de Resultados Parciais*

Após a execução dos métodos de treino, validação e teste são gerados e armazenados os seguintes resultados parciais: o valor da acurácia, a classificação da amostra fornecida pela rede, a classificação real da amostra e a *loss* do processo. Os valores serão utilizados posteriormente para o cálculo das métricas previamente descritas no Capítulo 3, de forma que seja possível avaliar o desempenho das redes e compará-los entre eles.

## 4 RESULTADOS E DISCUSSÃO

Neste capítulo serão apresentados os resultados obtidos através do experimento, separados em sessões, e ao final serão apresentados a comparação entre os melhores. Para a avaliação dos resultados considera-se principalmente o conjunto de especificidade e sensibilidade, pois entende-se que com alta sensibilidade o classificador é eficaz para identificar os casos em que o paciente esteja doente quando ele realmente está, e a especificidade os casos em que o paciente não esteja doente quando ele realmente não está.

Cada sessão contemplará uma breve análise dos valores, seguida de uma tabela que apresente a média dos resultados obtidos no experimento ao decorrer de quatro execuções, a matriz de confusão correspondente e um gráfico com os valores da *loss* da configuração do modelo.

### 4.1 VGG16

Para a VGG16 obtiveram-se resultados não-satisfatórios com ambas as taxas de aprendizado, pois com menos de 0,5 para a acurácia do classificador pode-se presumir que ele seja mais propício a errar a classificação de uma amostra do que acertar, sendo pior do que uma classificação feita aleatoriamente. Conforme as Tabelas 4.1 e 4.2, pode-se perceber que os resultados da especificidade são os piores e mais esparsos, com o maior desvio padrão entre as métricas avaliadas, o que indica fortemente problemas com a distinção das classes positivas de negativas. Ainda, percebe-se que houveram problemas com a *loss* dos modelos, pois existem picos com valores altíssimos e indesejados ocorrendo com as duas taxas de aprendizado, como pode-se verificar nas Figuras 13 e 16.

Ainda, conforme a Tabela 4.3, percebe-se que todos os resultados obtidos com a taxa de aprendizado de 0,001 foram todos superiores.

### 4.2 VGG19

Quanto à VGG19, os resultados obtidos foram melhores comparados aos da VGG16. Também abaixo de 0,5, no entanto, os resultados desta rede apresentaram melhor consistência em relação à anterior. Os resultados na Tabela ?? mostram que a taxa de aprendizado de 0,0001 se mostrou superior, tendendo a se aproximar de 0,5. Por sua vez, a *loss* destas redes já se mostram fixas em intervalos abaixo de 1, e o valor dos intervalos que as respectivas *losses* atingem são proporcionais à qualidade dos resultados das redes, percebendo-se que com a taxa em 0,0001 o intervalo atinge picos menores em relação a taxa em 0,001 (Figuras 18 e 19). É perceptível também que os grandes desvios padrão que a VGG16 apresentou foram atenuados nos resultados deste modelo.

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.429167	0.408333	0.419589	0.412830	0.441667
Desvio Padrão	0.043833	0.087665	0.050733	0.068209	0.050000
Valor Mínimo	0.366667	0.333333	0.357143	0.344828	0.400000
Valor Máximo	0.466667	0.466667	0.466667	0.476190	0.500000

Tabela 4.1: Resultados da média de quatro dias de execução da rede VGG16 com taxa de aprendizado 0,001

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.404167	0.383333	0.398500	0.389677	0.408333
Desvio Padrão	0.047871	0.079349	0.054599	0.063871	0.056928
Valor Mínimo	0.350000	0.300000	0.333333	0.315790	0.333333
Valor Máximo	0.466667	0.466667	0.464286	0.448276	0.466667

Tabela 4.2: Resultados da média de quatro dias de execução da rede VGG16 com taxa de aprendizado 0,0001

Métricas	LR = 0,001	LR = 0,0001
Acuracia	0,429167	0,404167
Precisão	0,441667	0,408333
Especificidade	0,408333	0,383333
Sensibilidade	0,419589	0,398500
F1Score	0,412830	0,389677

Tabela 4.3: Resultados agregados do desempenho da VGG16

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.458333	0.441667	0.446759	0.443548	0.475000
Desvio Padrão	0.206604	0.239405	0.202913	0.220895	0.177169
Valor Mínimo	0.333333	0.300000	0.333333	0.315790	0.333333
Valor Máximo	0.766667	0.800000	0.750000	0.774194	0.733333

Tabela 4.4: Resultados da média de quatro dias de execução da rede VGG19 com taxa de aprendizado 0,001

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.466667	0.433333	0.470960	0.450525	0.500000
Desvio Padrão	0.238436	0.224433	0.256771	0.239455	0.262467
Valor Mínimo	0.283333	0.300000	0.290323	0.295082	0.266667
Valor Máximo	0.816667	0.766667	0.851852	0.807018	0.866667

Tabela 4.5: Resultados da média de quatro dias de execução da rede VGG19 com taxa de aprendizado 0,0001

Métricas	LR = 0,001	LR = 0,0001
Acuracia	0,458333	0,466667
Precisão	0,475000	0,500000
Especificidade	0,441667	0,433333
Sensibilidade	0,446759	0,470960
F1Score	0,443548	0,450525

Tabela 4.6: Resultados agregados do desempenho da VGG19

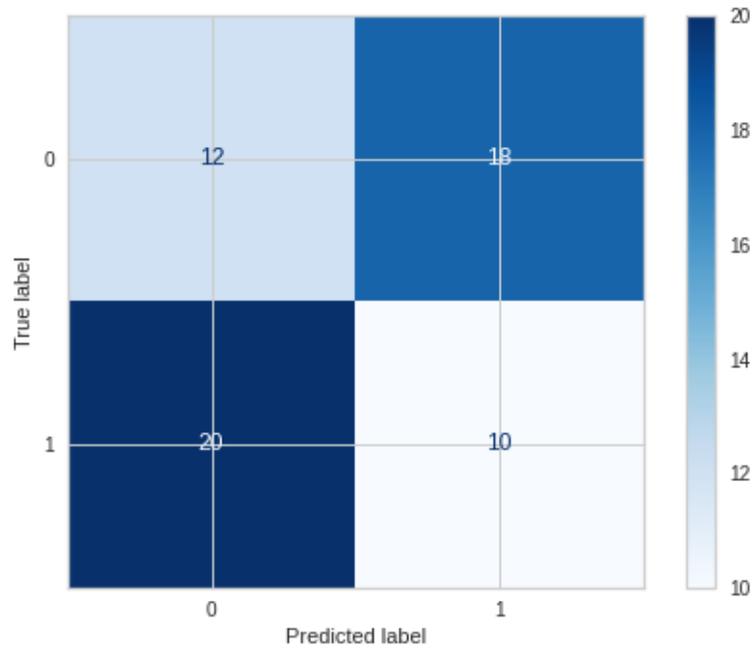


Figura 4.1: Matriz de confusão da vgg16 com learning rate = 0,001.

Fonte: Os autores

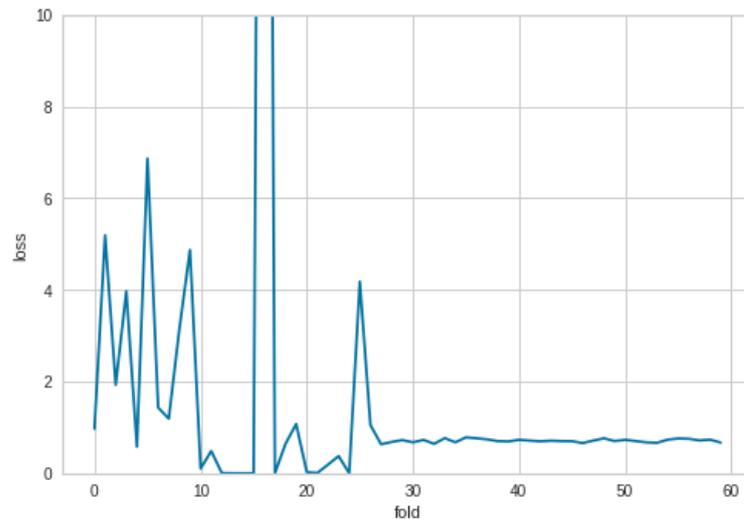


Figura 4.2: Média da loss da vgg16 com learning rate = 0,001.

Fonte: Os autores

### 4.3 INCEPTIONV3

Para a InceptionV3 os resultados passam a ser superiores às redes VGG16 e VGG19. Todos os valores mínimos das métricas da InceptionV3 (tabela 4.7 e 4.8) se apresentam superiores aos valores das médias das duas primeiras redes apresentadas. Com a taxa de aprendizado em 0,001 foram apresentados os melhores resultados, à parte da *loss*. Os problemas continuam por ser a alta variabilidade dos valores da *loss*, com faixas constantes tendendo a 0 mas ainda com picos recorrentes (Figuras 25 e 28). Com ambas as taxas de aprendizado, percebe-se que a sensibilidade das redes foram superiores a especificidade, o que indica maior sucesso para a

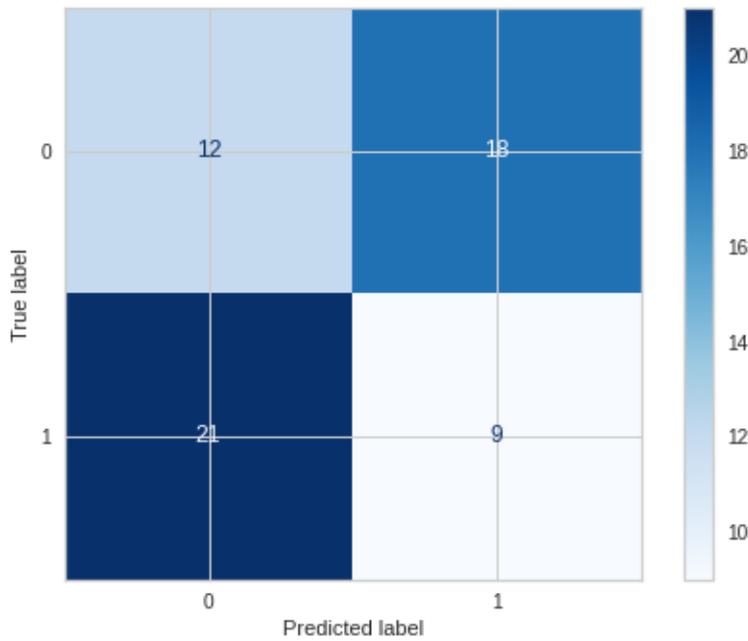


Figura 4.3: Matriz de confusão da vgg16 com learning rate = 0,0001.

Fonte: Os autores

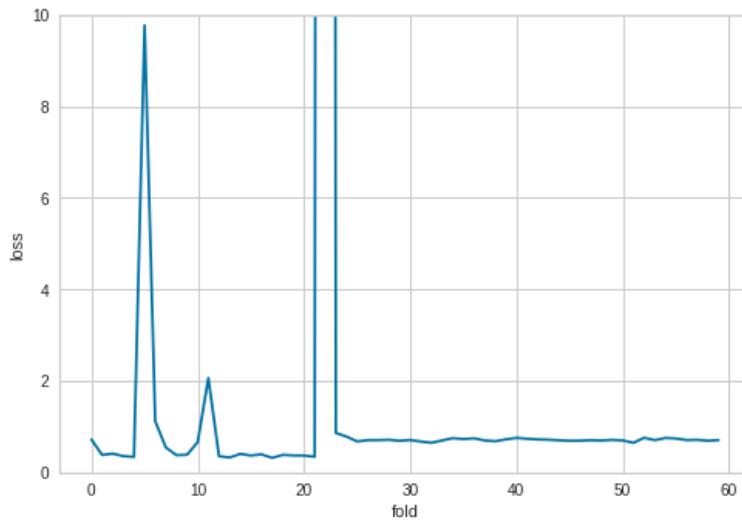


Figura 4.4: Média da loss da vgg16 com learning rate = 0,0001.

Fonte: Os autores

classificação de amostras da classe patológica (evidenciado pela matriz de confusão constada na Figura 29).

inceptionV3 0,0001 valores proximos das outras metricas, mas a acuracia é extremamente consistente, tem um desvio padrão bem baixo

#### 4.4 RESNET50V2

A ResNet50V2 apresentou bons resultados, similarmente à InceptionV3. O uso da taxa de aprendizado em 0,0001 apresentou resultados levemente inferiores à sua contraparte, com a taxa utilizada em 0,001, que por sua vez apresentou superioridade da especificidade em

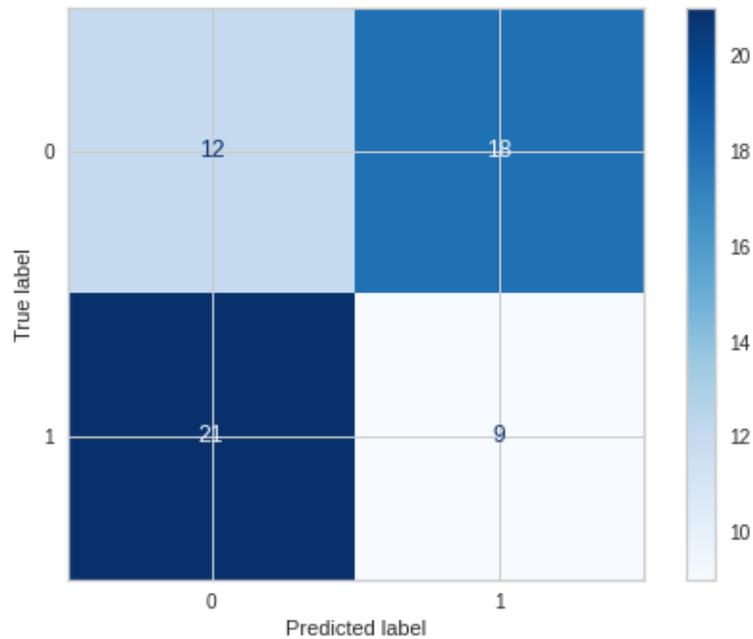


Figura 4.5: Matriz de confusão da vgg19 com learning rate = 0,001.

Fonte: Os autores

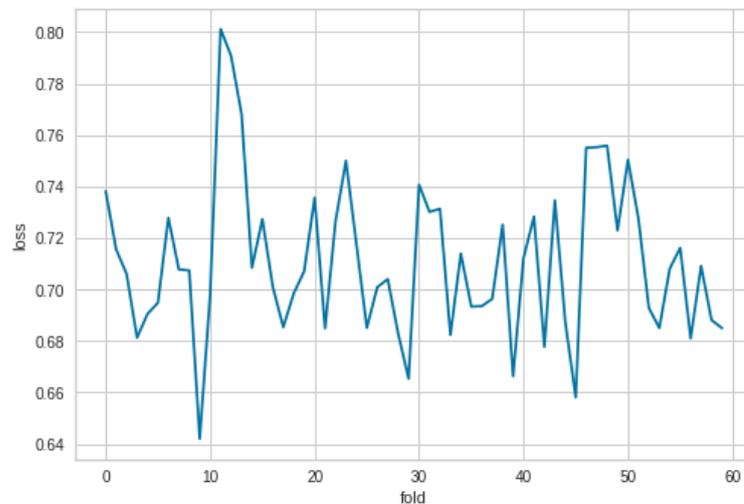


Figura 4.6: Média da loss da da vgg19 com learning rate = 0,001.

Fonte: Os autores

relação as demais métricas, demonstrando maior facilidade para a classificação das amostras pertencentes a classe normal. Os problemas com a *loss* ainda se mostram presentes, no entanto. Ademais, não pode-se concluir que a ResNet50V2 seja estritamente melhor que a InceptionV3.

métricas também bem consistentes, mas com Sensibilidade e precisão um pouco mais esparsadas, um pouco pior em achar classes que possuem doença

#### 4.5 ANÁLISE DOS RESULTADOS

Por fim, considera-se que a InceptionV3 com taxa de aprendizado em 0,001 seja a rede mais adequada para o problema, seguida da ResNet50V2 com taxa de aprendizado de 0,001. Além de ter a maior acurácia, a InceptionV3 apresenta consequentemente balanço entre

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.837500	0.841667	0.837331	0.837382	0.850000
Desvio Padrão	0.034359	0.073912	0.044977	0.039351	0.057735
Valor Mínimo	0.800000	0.733333	0.794118	0.785714	0.766667
Valor Máximo	0.883333	0.900000	0.896552	0.881356	0.900000

Tabela 4.7: Resultados da média de quatro dias de execução da rede InceptionV3 com taxa de aprendizado 0,001

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.825000	0.858333	0.812544	0.829600	0.841667
Desvio Padrão	0.021517	0.095743	0.060196	0.025476	0.056928
Valor Mínimo	0.800000	0.733333	0.736842	0.800000	0.766667
Valor Máximo	0.850000	0.933333	0.880000	0.861538	0.900000

Tabela 4.8: Resultados da média de quatro dias de execução da rede InceptionV3 com taxa de aprendizado 0,0001

Métricas	LR = 0,001	LR = 0,0001
Acuracia	0,837500	0,816667
Precisão	0,850000	0,841667
Especificidade	0,841667	0,833333
Sensibilidade	0,837331	0,816725
F1Score	0,837382	0,818954

Tabela 4.9: Resultados agregados do desempenho da InceptionV3

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.816667	0.833333	0.817275	0.824506	0.808333
Desvio Padrão	0.075768	0.800000	0.705882	0.750000	0.666667
Valor Mínimo	0.716667	0.800000	0.705882	0.750000	0.666667
Valor Máximo	0.883333	0.866667	0.896552	0.881356	0.900000

Tabela 4.10: Resultados da média de quatro dias de execução da rede ResNet50V2 com taxa de aprendizado 0,001

	Acurácia	Especificidade	Sensibilidade	F1Score	Precisão
Média	0.800000	0.825000	0.795002	0.808948	0.783333
Desvio Padrão	0.054433	0.050000	0.075431	0.057231	0.088192
Valor Mínimo	0.733333	0.766667	0.718750	0.741935	0.700000
Valor Máximo	0.866667	0.866667	0.896552	0.881356	0.900000

Tabela 4.11: Resultados da média de quatro dias de execução da rede InceptionV3 com taxa de aprendizado 0,0001

Métricas	LR = 0,001	LR = 0,0001
Acuracia	0,816667	0,800000
Precisão	0,808333	0,783333
Especificidade	0,833333	0,825000
Sensibilidade	0,817275	0,795002
F1Score	0,824506	0,808948

Tabela 4.12: Resultados agregados do desempenho da ResNet50V2

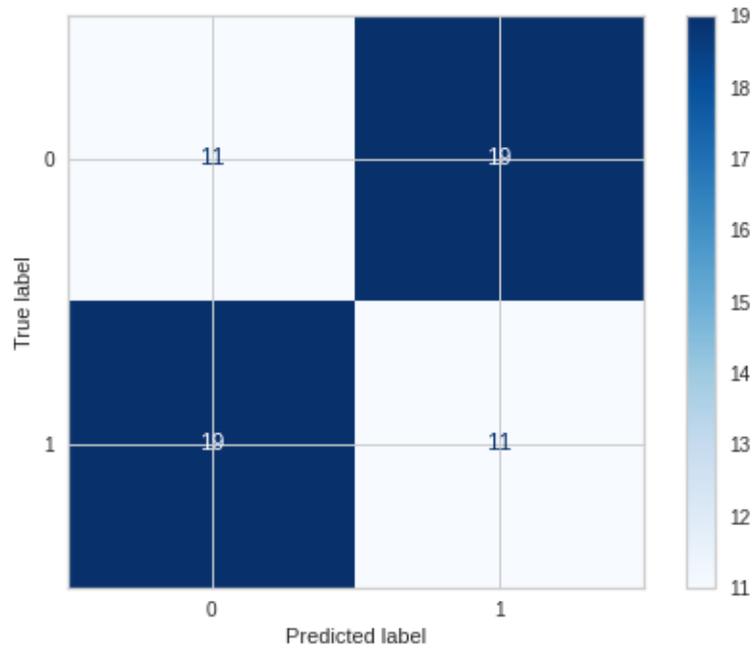


Figura 4.7: Matriz de confusão da vgg19 com learning rate = 0,0001.

Fonte: Os autores

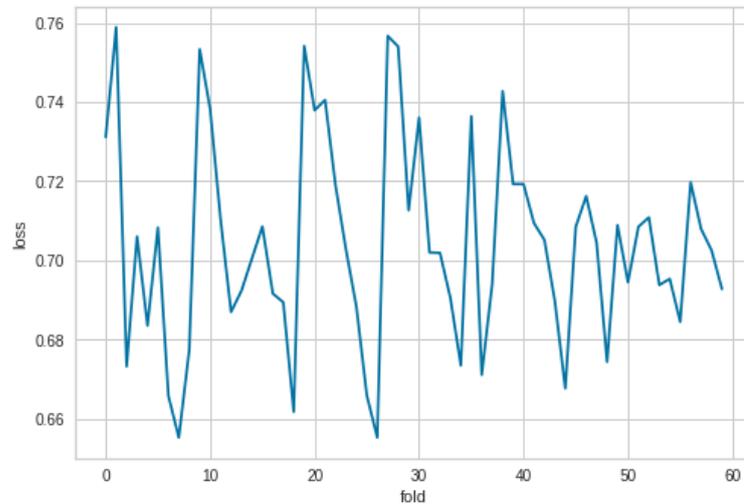


Figura 4.8: Média da loss da da vgg19 com learning rate = 0,0001.

Fonte: Os autores

os valores de especificidade e sensibilidade, o que deve ser fortemente valorizado no contexto de auxílio ao diagnóstico sem que um foco tenha sido especificado, pelo veterinário, para o problema. Por possuir também as melhores precisão e score f1, percebe-se que os resultados positivos apresentados por ela são inteiramente superiores às demais, podendo demonstrar sua superioridade.

As redes VGG16 e VGG19 por mais que tivessem apresentado os piores resultados, acabaram por serem mais consistentes devida a uma variação não tão grande da *loss*. Em contrapartida, tanto a inceptionV3 quanto a ResNet50V2 tiveram inconstancias muito grandes nos valores de suas *losses*, podendo significar que tanto no acerto quanto no erro elas têm uma certeza muito grande de do resultado fornecido, ou seja, acerta por muito e erra igualmente por muito. Acredita-se que o mal comportamento tenha sido causado por uma mudança no processo

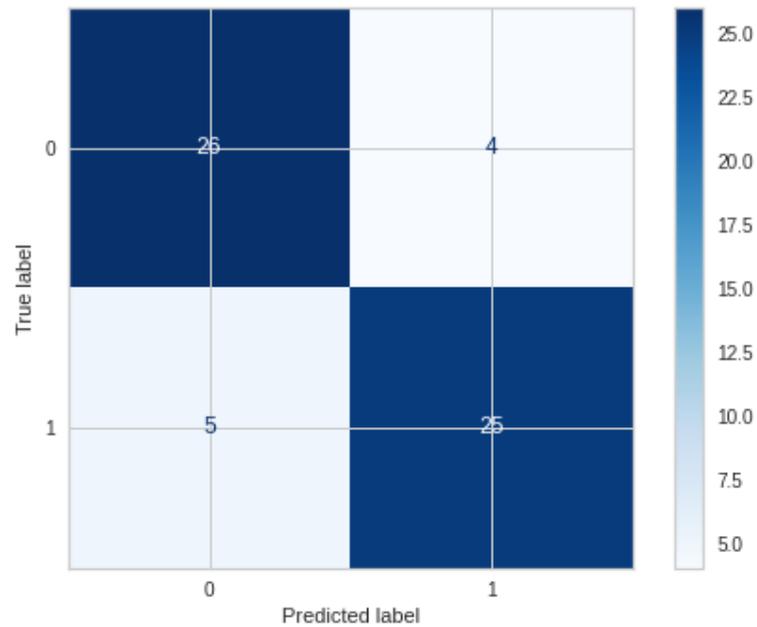


Figura 4.9: Matriz de confusão da inceptionV3 com learning rate = 0,001.  
Fonte: Os autores

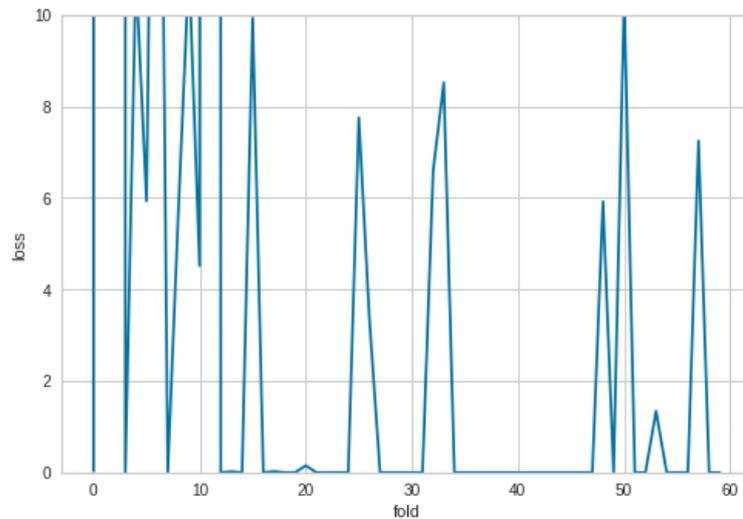


Figura 4.10: Média da loss da da inceptionV3 com learning rate = 0,001.  
Fonte: Os autores

de salvar e atualizar os pesos das redes no processo de treino, mas a correção dos resultados com a versão anterior não é possível devido às limitações proporcionadas pela plataforma Google Colab, que exige por volta do equivalente à assinatura do plano Pro+ uma segunda vez.

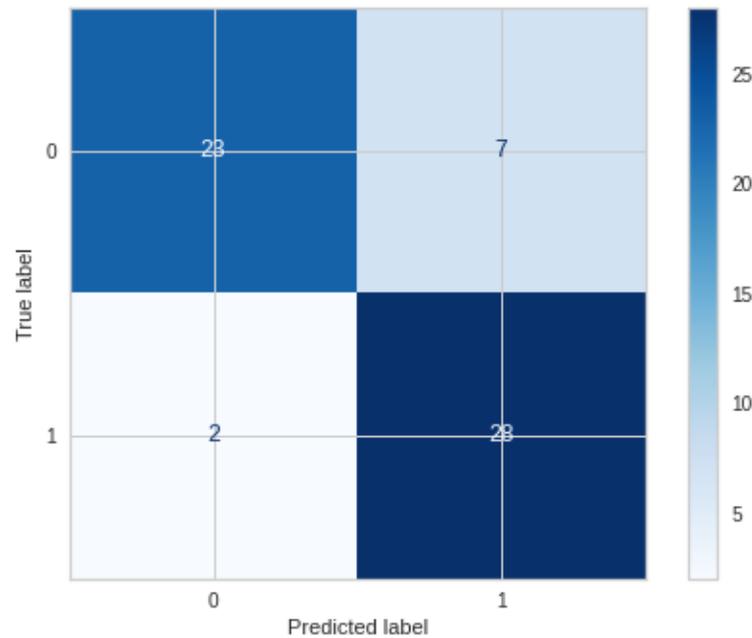


Figura 4.11: Matriz de confusão da inceptionV3 com learning rate = 0,0001.  
Fonte: Os autores

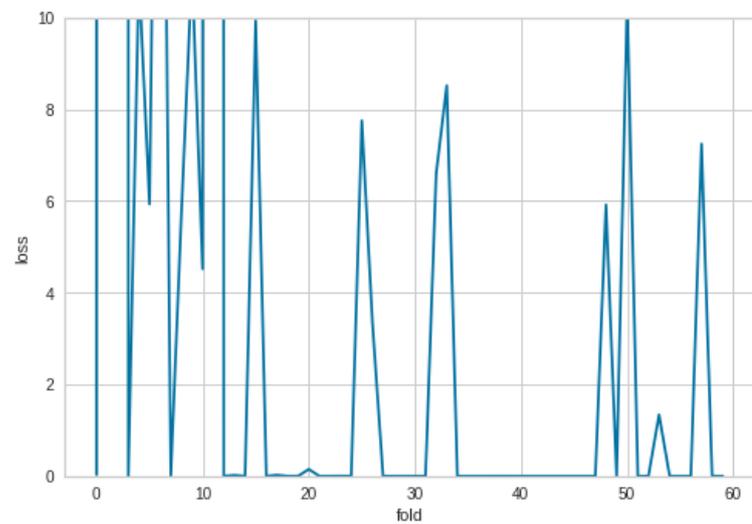


Figura 4.12: Média da loss da da inceptionV3 com learning rate = 0,0001.  
Fonte: Os autores

Redes	LR	Acurácia	Precisão	Especificidade	Sensibilidade	F1Score
VGG16	0,001	0,429167	0,441667	0,408333	0,419589	0,412830
VGG19	0,0001	0,466667	0,500000	0,433333	0,470960	0,450525
InceptionV3	0,001	0,837500	0,850000	0,841667	0,837331	0,837382
ResNet50V2	0,001	0,816667	0,808333	0,833333	0,817275	0,824506

Tabela 4.13: Agrupamento dos melhores resultados por rede

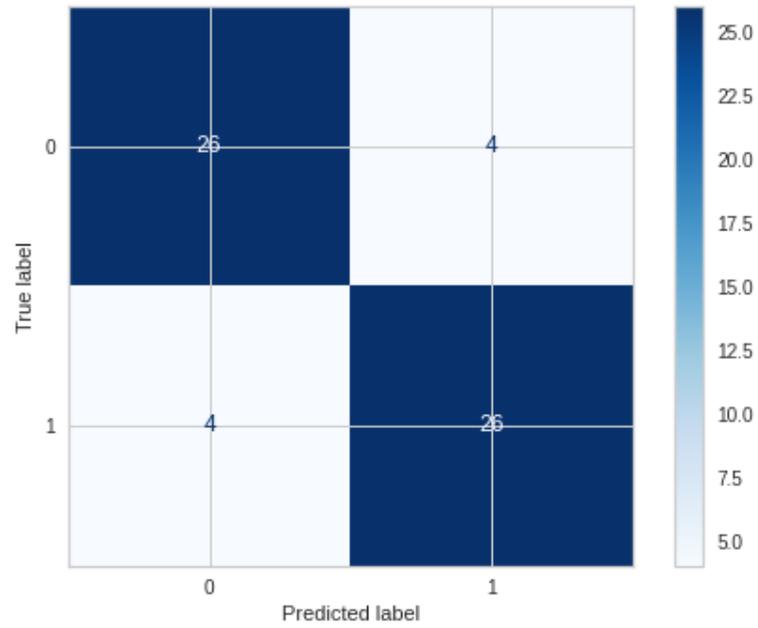


Figura 4.13: Matriz de confusão da resNet50V2 com learning rate = 0,001.  
Fonte: Os autores

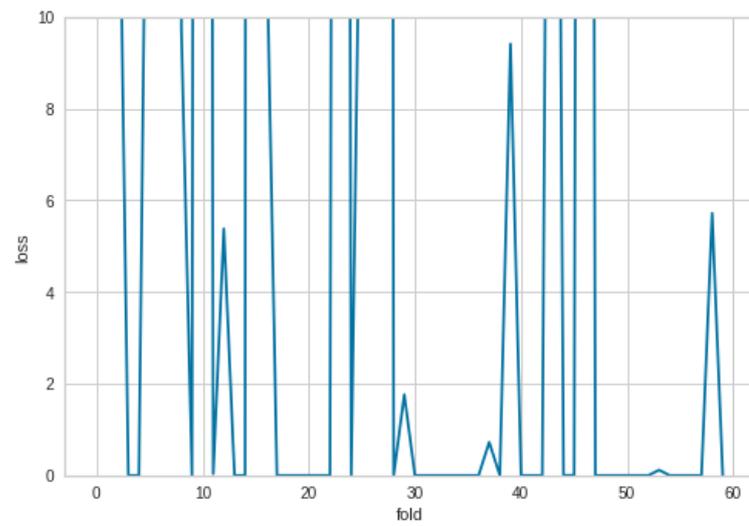


Figura 4.14: Média da loss da da resNet50V2 com learning rate = 0,001.  
Fonte: Os autores

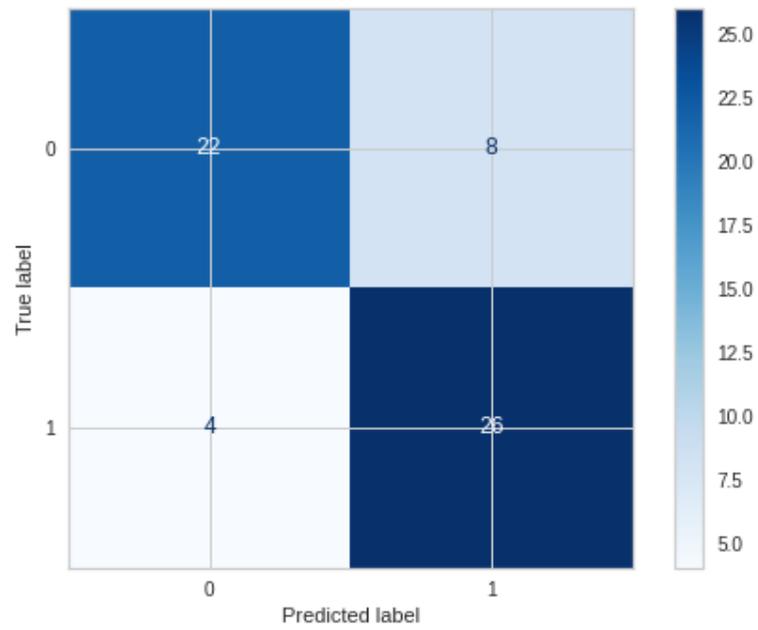


Figura 4.15: Matriz de confusão da resNet50V2 com learning rate = 0,0001.  
Fonte: Os autores

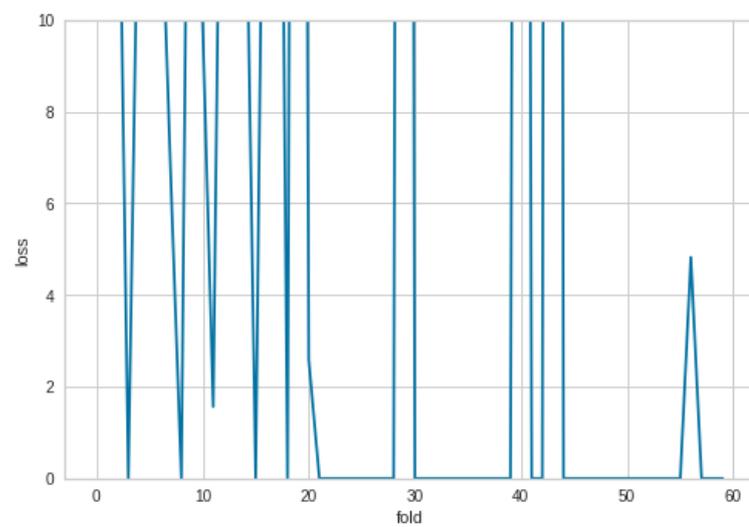


Figura 4.16: Média da loss da da resNet50V2 com learning rate = 0,0001.  
Fonte: Os autores

## 5 CONCLUSÃO E SUGESTÕES PARA TRABALHOS FUTUROS

No decorrer do desenvolvimento deste trabalho algumas questões quanto aos resultados dos experimentos foram levantadas e algumas dificuldades foram enfrentadas. Ao levar em consideração a influência desses dois fatores, que serão descritos posteriormente, os autores possuem algumas sugestões para a melhoria

Uma primeira dificuldade foi a limitação do ambiente de execução do experimento. A execução dos experimentos inicialmente esboçados, com a perspectiva da aquisição de um *dataset* composto por mais amostras e a aplicação de outras técnicas para o treino, teste e validação das redes, não possibilitaram a descoberta das limitações diárias de quota e do tempo necessário para o processamento do experimento na plataforma, pois não apresentaram necessidade e condições de consumir os mesmos recursos do experimento em sua versão final. Sendo assim, uma primeira sugestão para um trabalho futuro é a adoção de um ambiente menos limitante, para que seja possível gerar mais resultados sem depender dos limites diários impostos pelo Colab.

Outro fator que influenciou na demora para a descoberta das limitações de plataforma foram as limitações de aquisição do *dataset*. O processo de rotulação das imagens pelos profissionais capacitados para a função e a adequação das amostras para possibilitar que sejam utilizadas no experimento (anonimização de informações sensíveis e conversão de Dicom para png) foram custosos e consumiram tempo. O *dataset* utilizado no trabalho resulta das amostras que estavam dentro da possibilidade de serem disponibilizadas no período do desenvolvimento do trabalho, então uma outra melhoria para tal é o uso de outra versão do *dataset*, que contenha mais imagens e permita o uso de outras técnicas para o treino das redes neurais propostas. Idealmente, composto por imagens também providas de um laboratório distinto daquele das imagens originalmente utilizadas.

Quanto aos resultados, algumas questões foram levantadas, a começar pela variação dos hiperparâmetros. Os resultados apresentados neste trabalho levam apenas em consideração a variação na taxa de aprendizado devido às limitações previamente mencionadas, mas considera-se interessante a possibilidade de avaliar o desempenho das quatro redes utilizadas com variações mais drásticas em suas configurações. Quanto à análise dos resultados com maior variação dos hiperparâmetros, considera-se também interessante a aplicação do método *Grid Search*, para que a partir dos melhores resultados para cada rede a comparação entre elas possa ser feita.

Por fim, uma outra análise dos resultados que considera-se interessante é relativa às características extraídas das amostras. De forma similar ao processo de segmentação de uma região em uma imagem utilizando aprendizado profundo, os autores sugerem a criação de um mapa de calor, a partir das características extraídas, na respectiva amostra fornecida como entrada para a rede. Localizando a principal região de que as características estejam sendo extraídas é possível pensar em etapas de pré-processamento das amostras para que a classificação seja facilitada e obtenha melhores resultados.

Conforme previamente exposto, o problema de aumento de átrio é frequente para os casos de problemas cardíacos em cães, e sua identificação adianta o tratamento dos pacientes. No entanto, o diagnóstico do sintoma é custoso, podendo ser considerado um gargalo do processo que pode ser feito pela adoção de ferramentas de diagnóstico auxiliado por computador. A ferramenta pode proporcionar contribuições variando de melhorias no nível de certeza do diagnóstico do profissional e idealmente chegando ao ponto de reduzir o volume de exames que necessitem chegar às mãos do profissional por filtrar os casos negativos à presença do sintoma em uma análise prévia dos dados dos pacientes.

Neste trabalho, através de experimentação, obteve-se resultados indicando a possibilidade de a rede de aprendizado profundo InceptionV3 ser a mais propícia a ser utilizada no desenvolvimento de uma ferramenta de auxílio ao diagnóstico no processamento de imagens radiográficas, pois suas classificações foram indicadas de terem certeza tanto nos casos de presença do sintoma quanto nos casos de ausência. Um complicador, no entanto, é o comportamento da métrica *loss* da rede, que levanta questionamentos a serem respondidos com a execução de novos experimentos em plataformas distintas.

## REFERÊNCIAS

- Atkins, C., Bonagura, J., Ettinger, S., Fox, P., Gordon, S., Haggstrom, J., Hamlin, R., Keene, B., Luis-Fuentes, V. e Stepien, R. (2009). Guidelines for the diagnosis and treatment of canine chronic valvular heart disease. *Journal of veterinary internal medicine*, 23(6):1142–1150.
- Awan, A. A. (2022). A complete guide to data augmentation. Acessado em 11/02/2023.
- Azevedo-Marques, P. M. d. (2001). Diagnóstico auxiliado por computador na radiologia. *Radiologia Brasileira*, 34:285–293.
- Banzato, T., Wodzinski, M., Burti, S., Osti, V. L., Rossoni, V., Atzori, M. e Zotti, A. (2021a). Automatic classification of canine thoracic radiographs using deep learning. *Scientific Reports*, 11(1):3964.
- Banzato, T., Wodzinski, M., Tauceri, F., Donà, C., Scavazza, F., Müller, H. e Zotti, A. (2021b). An ai-based algorithm for the automatic classification of thoracic radiographs in cats. *Frontiers in veterinary science*, 8:731936.
- Das, A. K., Ghosh, S., Thunder, S., Dutta, R., Agarwal, S. e Chakrabarti, A. (2021). Automatic covid-19 detection from x-ray images using ensemble learning with convolutional neural network. *Pattern Analysis and Applications*, 24:1111–1124.
- Doi, K. (2007). Computer-aided diagnosis in medical imaging: Historical review, current status and future potential. *Computerized Medical Imaging and Graphics*, 31(4):198–211. Computer-aided Diagnosis (CAD) and Image-guided Decision Support.
- e Silva, S. R. e Schimidt, F. (2016). Redução de variáveis de entrada de redes neurais artificiais a partir de dados de análise de componentes principais na modelagem de oxigênio dissolvido. *Quim. Nova*, 39(3):273–278.
- He, K., Zhang, X., Ren, S. e Sun, J. (2016). Deep residual learning for image recognition. Em *Proceedings of the IEEE conference on computer vision and pattern recognition*, páginas 770–778.
- Khan, I. U. e Aslam, N. (2020). A deep-learning-based framework for automated diagnosis of covid-19 using x-ray images. *Information*, 11(9):419.
- Li, S., Wang, Z., Visser, L. C., Wisner, E. R. e Cheng, H. (2020). Pilot study: application of artificial intelligence for detecting left atrial enlargement on canine thoracic radiographs. *Veterinary radiology & ultrasound*, 61(6):611–618.
- Lin, M., Chen, Q. e Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.
- Nogare, D. (2020). Performance de machine learning – matriz de confusão. Acessado em 17/02/2023.
- Perez, L. e Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.

- Simonyan, K. e Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J. e Wojna, Z. (2016). Rethinking the inception architecture for computer vision. Em *Proceedings of the IEEE conference on computer vision and pattern recognition*, páginas 2818–2826.
- Tech, D. (2022). O pacote caret – linguagem r. Acessado em 30/03/2023.