

MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM COMPUTACIONAL

**APLICAÇÃO DE REDES NEURAIS CONVOLUCIONAIS PARA
AUXILIAR O DIAGNÓSTICO DE PNEUMONIA VIRAL E
BACTERIOLÓGICA EM IMAGENS DE RAIO-X**

por

Felipe dos Santos Vieira

Dissertação para obtenção do Título de
Mestre em Modelagem Computacional

Rio Grande, junho, 2025

Felipe dos Santos Vieira

“Aplicação de Redes Neurais Convolucionais para Auxiliar o Diagnóstico de Pneumonia Viral e Bacteriológica em Imagens de Raio-x”

Dissertação apresentada ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal do Rio Grande - FURG, como requisito parcial para obtenção do Grau de Mestre. Área de concentração: Modelagem Computacional.

Aprovada em: 28/02/2025

BANCA EXAMINADORA

Documento assinado digitalmente
 **DINALVA AIRES DE SALES**
Data: 28/02/2025 11:41:42-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Dinalva Aires de Sales (Orientadora – FURG)

Documento assinado digitalmente
 **JEAN MARCEL DE ALMEIDA ESPINOZA**
Data: 28/02/2025 15:42:52-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Jean Marcel de Almeida Espinoza (Coorientador – IFSC)

Documento assinado digitalmente
 **SAMUEL DA SILVA FEITOSA**
Data: 28/02/2025 11:45:55-0300
Verifique em <https://validar.iti.gov.br>

Prof. Dr. Samuel da Silva Feitosa (UFFS)

Documento assinado digitalmente
 **DIANA FRANCISCA ADAMATTI**
Data: 28/02/2025 18:04:05-0300
Verifique em <https://validar.iti.gov.br>

Profa. Dra. Diana Francisca Adamatti (FURG)

Documento assinado digitalmente
 **ALEXANDRE DA SILVA CAPIVERDE**
Data: 28/02/2025 16:24:50-0300
Verifique em <https://validar.iti.gov.br>

Alexandre Da Silva Capaverde (HU - Doutor Miguel Riet Correa Júnior)

Rio Grande - RS
2025

Ficha Catalográfica

V658a Vieira, Felipe dos Santos.

Aplicação de redes neurais convolucionais para auxiliar o diagnóstico de pneumonia viral e bacteriológica em imagens de raio-x / Felipe dos Santos Vieira. – 2025.

89 f.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG, Programa de Pós-Graduação em Modelagem Computacional, Rio Grande/RS, 2025.

Orientadora: Dra. Dinalva Aires de Sales.

Coorientador: Dr. Jean Marcel de Almeida Espinoza.

1. Aprendizado de Máquina 2. Diagnóstico Médico 3. Redes Neurais Convolucionais 4. Raios-x 5. VGG16 I. Sales, Dinalva Aires de II. Espinoza, Jean Marcel de Almeida III. Título.

CDU 004:616.24

Catálogo na Fonte: Bibliotecário José Paulo dos Santos CRB 10/2344

AGRADECIMENTOS

A jornada para a conclusão deste trabalho foi marcada por desafios, aprendizados e pelo apoio inestimável de diversas pessoas e instituições, às quais expresso minha profunda gratidão.

Agradeço a Deus pela força e inspiração ao longo deste caminho. À minha família, pelo amor incondicional, paciência e apoio em todos os momentos. Em especial, à minha avó, à minha irmã e à minha namorada Luana, que sempre acreditaram no meu potencial e me incentivaram a buscar meus objetivos.

Meu sincero agradecimento a Dinalva Aires de Sales, por sua orientação, paciência e dedicação. Sua experiência e conselhos foram fundamentais para o desenvolvimento desta dissertação e para o meu crescimento acadêmico e profissional.

Sou grato à Universidade Federal do Rio Gande - Furg por proporcionar os recursos e o ambiente necessários para a realização deste trabalho. Agradeço também aos colegas de laboratório/curso, pela troca de conhecimentos, apoio mútuo e pelos momentos de descontração que tornaram essa caminhada mais leve.

Aos amigos que estiveram ao meu lado, em especial ao Victor oferecendo palavras de incentivo e momentos de descontração, especialmente nos períodos mais desafiadores.

Agradeço à agência de fomento, CAPES, pelo apoio financeiro e pela confiança em minha pesquisa.

Por fim, minha gratidão a todos que, de alguma forma, contribuíram para que este trabalho se tornasse realidade. Cada gesto de apoio foi indispensável para a conclusão deste projeto.

RESUMO

O diagnóstico preciso de pneumonia com base em imagens médicas é fundamental para a melhoria da saúde pública. Este trabalho propõe um modelo de classificação de pneumonia como um auxílio ao diagnóstico médico, utilizando redes neurais convolucionais, especificamente a arquitetura baseada no modelo VGG16, combinada com camadas densas adicionais. O modelo foi treinado e avaliado em um conjunto de imagens de raios-X da região torácica de crianças, jovens e adultos, utilizando o algoritmo de otimização SGD e a função de perda *categorical crossentropy*. Técnicas como *early stopping* e *model checkpoint* foram empregadas para evitar sobreajuste. Os resultados alcançaram uma acurácia de 84%, sensibilidade de 83,0% e especificidade de 83,8%, com uma área sob a curva ROC (AUC) de 83,9% para a classificação de pneumonia versus normal. Apesar de os valores serem ligeiramente inferiores aos de estudos, os resultados são competitivos e demonstram a eficácia do modelo em condições reais. Além disso, o trabalho destaca a importância de utilizar conjuntos de dados públicos para fomentar a inovação e o desenvolvimento de algoritmos de aprendizado de máquina na área médica. Por fim, foi realizada uma análise comparativa com outros estudos da literatura, evidenciando as contribuições e limitações do modelo proposto, bem como possíveis direções para pesquisas futuras.

Palavras-chave: Aprendizado de Máquina; Diagnóstico Médico; Redes Neurais Convolucionais; Raios-x; VGG16.

ABSTRACT

Accurate diagnosis of pneumonia based on medical images is essential for improving public health. This work proposes a pneumonia classification model as an aid to medical diagnosis, using convolutional neural networks, specifically the architecture based on the VGG16 model, combined with additional dense layers. The model was trained and evaluated on a set of chest X-ray images of children, adolescents and adults, using the SGD optimization algorithm and the categorical crossentropy loss function. Techniques such as early stopping and model checkpoint were employed to avoid overfitting. The results achieved an accuracy of 84%, sensitivity of 83.0% and specificity of 83.8%, with an area under the ROC curve (AUC) of 83.9% for the classification of pneumonia versus normal. Although the values are slightly lower than those of studies, the results are competitive and demonstrate the effectiveness of the model in real conditions. Furthermore, the work highlights the importance of using public datasets to foster innovation and the development of machine learning algorithms in the medical field. Finally, a comparative analysis was carried out with other studies in the literature, highlighting the contributions and limitations of the proposed model, as well as possible directions for future research.

Keywords: Machine Learning; Medical Diagnosis; Convolutional Neural Networks; X-rays; VGG16.

ÍNDICE

1	INTRODUÇÃO	14
1.1	Introdução	14
1.2	Motivação	15
1.3	Objetivos	16
1.3.1	Objetivo geral	16
1.3.2	Objetivos específicos	16
1.4	Estrutura do trabalho	16
2	EMBASAMENTO TEÓRICO	18
2.1	Diagnóstico médico	18
2.2	Pneumonia	19
2.3	Laudo médico	21
2.4	Inteligencia artificial no diagnóstico médico	23
2.5	Redes neurais artificiais	23
2.6	Redes neurais de classificação binárias	26
2.7	Redes neurais de classificação multiclasse	26
2.8	Aspectos Matemáticos e Computacionais das Redes Neurais Convolucionais (CNNs)	26
2.9	Camadas Convolucionais	26
2.10	Camadas de Pooling	27
2.11	Camadas Totalmente Conectadas	27
2.12	Treinamento da Rede Neural Convolucional	28
2.12.1	Forward Pass (Propagação para Frente)	28
2.12.2	Cálculo da Função de Perda	28
2.12.3	Backpropagation e Atualização dos Pesos	29
2.13	Otimizações no Treinamento	29
2.14	Linguagem e pacotes	30
2.15	<i>Visual Geometry Group</i> (VGG)	31
2.16	Trabalhos relacionados	35
3	METODOLOGIA	41
3.1	<i>Dataset</i>	41
3.2	Seleção de amostras	41
3.3	Treinamento da rede neural	42
3.4	Componentes da Rede Neural	46
3.5	Métricas de avaliação do modelo	48
3.5.1	Métricas Derivadas da Matriz de Confusão	49
4	RESULTADOS E DISCUSSÕES	54
4.1	<i>Dataset</i>	54
4.2	Treinamento e Validação do Modelo	55
4.3	Matriz de Confusão	60

4.4	Análise das Métricas: Precisão, Recall e F1-score	65
4.5	Curva ROC	67
4.6	Cálculo da AUC para o modelo	70
4.7	Cálculo da Curva ROC e AUC para Múltiplas Classes	71
4.8	Cálculo da AUC para cada classe (One-vs-Rest)	72
4.9	Classe Normal (One-vs-Rest)	72
4.10	Classe Bactéria (One-vs-Rest)	72
4.11	Classe Vírus (One-vs-Rest)	72
4.12	Análise de Normal vs Pneumonia	73
4.13	Auxílio ao diagnostico	80
5	CONCLUSÃO	84
	REFERÊNCIAS	86

LISTA DE FIGURAS

1	Imagem de pneumonia bacteriana	22
2	Imagem de pneumonia viral	22
3	Neurônio Biológico	24
4	Unidade lógica de Limiar	24
5	Problema de classificação XOR	25
6	Rede Neural CNN	25
7	Exemplo de Estrutura Neural	33
8	Arquitetura do Modelo VGG-16.	33
9	Modelo <i>VGG-16</i> input dos dados.	34
10	Configurações do Modelo VGG-16.	35
11	Rede Neural Artificial de Tsiknakis	36
12	Rede Neural Artificial de Ibrahim	36
13	Rede Neural Artificial de Bhandary	37
14	Rede Neural Artificial de Rezaee	38
15	Rede Neural Artificial de Kermany	39
16	Rede Neural Artificial de Ayan	40
17	<i>Dataset</i>	42
18	Dados Parciais Gerados Pela Rede Neural	43
19	Gráfico de Perda	44
20	Dados Gerados Pelo Treinamento da Rede VGG	44
21	Fluxograma do Modelo de Treinamento da Rede Neural	45
22	Matriz genérica para classificação multiclasse	50
23	Exemplo da Separação das Classes.	54
24	Gráfico das Métricas de Treinamento do Modelo.	56
25	Gráfico das métricas de validação do modelo.	57
26	Gráfico da Acurácia e Perda do Modelo.	58
27	Acurácia do modelo de Enes AYAN.	59
28	Acurácia do modelo de Rachna Jain.	59
29	Acurácia do modelo de Daniel S. Kermany.	60
30	Matriz de Confusão	62
31	Matriz de Confusão de AYAN; ÜNVER (2019).	63
32	Matriz de Confusão de IBRAHIM et al. (2021).	64
33	Matriz de Confusão de TSIKNAKIS et al. (2020).	64
34	Matriz de Confusão Normalizada em Porcentagem.	65
35	Curva ROC	75
36	Curva ROC para as Classes Pelo Metodo <i>One vs Rest</i>	76
37	Curva ROC para Normal vs Pneumonia	76
38	Curva Roc de BHANDARY et al. (2020)	77

39	Curva Roc de KERMANY et al. (2018)	78
40	Resposta do modelo para classe normal.	81
41	Resposta do modelo para classe bactéria.	82
42	Resposta do modelo para classe vírus.	83

LISTA DE TABELAS

1	Matriz de Confusão.	49
2	Distribuição do Banco de Dados para o Modelo	54
3	Métricas de Treinamento	55
4	Métricas de Validação	55
5	Métricas de Teste	56
6	Comparação das Métricas de Acurácia e Perda	60
7	Matriz de Confusão	61
8	Resultados das Métricas	66
9	Comparação das Métricas no Teste do Modelo	67
10	Métricas de Desempenho das Classes	69
11	Pontos Organizados para Cálculo da AUC	71
12	Resultados da AUC para Cada Classe (One-vs-Rest)	73
13	Matriz de confusão para Normal vs Pneumonia.	73
14	Comparação das métricas no Teste do modelo	75
15	Comparação dos Valores de AUC em Modelos de Classificação de Pneumonia	79
16	Comparação dos Valores de Sensibilidade e Especificidade em Modelos de Classificação de Pneumonia	79

LISTA DE ABREVIATURAS

ADC	Conversor Analógico-Digital
AEC	Controle Automático de Exposição
AM	Aprendizado de Máquina
AutoML	Aprendizado de Máquina Automático
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CCD	Dispositivos de Carga Acoplados
CNN	Redes Neurais Convolucionais
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CNR	Relação Contraste Ruído
CT	Tomografias Computadorizadas
DNN	Rede Neural Profunda
DPOC	Doença Pulmonar que Obstrui as Vias Aéreas
DQE	Detetive Eficiência Quântica
eV	Elétron-volt
FID	Foco Distância Imagem
FN	False Negative
FOV	Field of View
FP	False Positive
FPR	Taxa de falsos positivos
FURG	Universidade Federal do Rio Grande
IA	Inteligência Artificial
IRA	Infecções Respiratórias Agudas
keV	Quiloelétron-volt
Kv	Quilovoltagem
LET	Transferência Linear de Energia
LTU	Unidade de limiar linear
mAs	Mili-Ampere por segundo
MeV	Megaelétron-volt
ML	Machine Learning
MLP	Perceptrons Multicamada
MP	Mycoplasma Pneumoniae
MTF	Função de Transferência de Modulação
NPS	Espectro de Potência de Ruído

OD	Densidade Óptica
OID	Objeto Distância Imagem
OMS	Organização Mundial de Saúde
PAC	Pneumonias Adquiridas na Comunidade
PACS	Picture Archiving and Communication System
ReLU	Rectified Linear Unit
RM	Ressonância Magnética
RNA	Redes Neurais Artificiais
RNN	Redes Neurais Recorrentes
SRAG	Síndrome Respiratória Aguda Grave
TLU	Unidade lógica de limiar
TN	True Negative
TP	True Positive
TPR	Taxa de verdadeiros positivos
TPU	Tensor Processing Units
VGG	Visual Geometry Group
XOR	Exclusive OR

1 INTRODUÇÃO

Neste capítulo é abordada a importância do tema da pesquisa apresentado neste trabalho. Esta introdução foi dividida em três partes, sendo elas: Introdução; motivação e objetivos do trabalho.

1.1 Introdução

Quando pensamos em uma patologia e como identificá-la, estamos nos referindo a um diagnóstico médico. A partir das visualizações e conhecimentos de anatomia e fisiologia humana, podemos identificar e diagnosticar se um segmento anatômico está ou não em suas funções normais. Uma das formas mais comuns é o diagnóstico por imagem, que utiliza os equipamentos de radiação ionizante e não ionizante para a produção de imagens para a identificação desses segmentos da anatomia e fisiologia humana (SILVA; PATROCÍNIO; SCHIABEL, 2019).

Uma das principais áreas para diagnóstico é o radiodiagnóstico ou diagnóstico por imagem, onde imagens produzidas por raios-x são utilizadas para o diagnóstico. O processamento e pós-processamento das imagens médicas é fundamental para um diagnóstico eficaz, e para que a determinada patologia não evolua sem o devido tratamento. Na atualidade, a identificação de patologias para o diagnóstico é realizada por profissionais de diversas áreas, sendo esses médicos, engenheiros, físicos, biólogos, entre outros. O diagnóstico ajuda a identificar inúmeras patologias, desde os estágios iniciais aos avançados, a identificação das patologias é comumente relatada em um laudo, um documento utilizado pelos profissionais da saúde para a elaboração de seus respectivos tratamentos (CURRY; DOWDEY; MURRY, 1990).

As imagens para serem laudadas passam por um processo chamado de processamento de imagem, onde as imagens são melhoradas computacionalmente, para facilitar a diferenciação da anatomia e fisiológico a serem estudados. Atualmente, enfrentamos um grande problema: a produção de imagens para diagnóstico é maior do que a quantidade de profissionais disponíveis para analisá-las, processá-las e laudá-las. A identificação e classificação de imagem por Inteligência Artificial (IA) é uma área em constante evolução, com aplicações em diversas áreas, como segurança, monitoramento e diferenciação de objetos. Em imagens médicas, essas diferenciações que buscamos são mais sutis, as imagens são produzidas em escalas de padrões de 255 tons de cinza (GONZALEZ; WOODS, 2010).

O processamento de imagens médicas utilizando IA é uma das principais inovações tecnológicas que revolucionaram o campo da medicina nos últimos anos, permitindo melhorias significativas no diagnóstico, monitoramento e tratamento de doenças. A combinação de algoritmos de IA com imagens médicas, incluindo os raios-x, tomografia computadorizada (CT) e ressonância magnética (RM), permite que os profissionais de saúde classifiquem possíveis casos de patologias de forma mais eficiente. A separação dos possíveis casos de positivos pode agilizar e facilitar o processo de diagnóstico, separando possíveis patologias em um banco de dados de imagens. Essa abordagem permite a automatização de tarefas que previamente demandavam

análises manuais extensivas, como detecção de tumores, segmentação de órgãos e testes de lesão, até mesmo identificação de padrões complexos em imagens médicas. A IA também pode prever resultados clínicos, possibilitar o tratamento personalizado em algumas situações (JAUHAR; THOMES, 2025).

No entanto, a implementação de soluções de processamento de imagens médicas baseadas em IA requer validação cuidadosa na prática clínica na verificação de patologias pré-separadas e classificadas, a regulamentação adequada para garantir a segurança e eficácia dos cuidados médicos. Portanto, a interseção entre processamento de imagens médicas e AI promete permanecer central para qualquer transformação digital na saúde, permitindo que os médicos forneçam cuidados aos pacientes com maior rapidez e precisão.

1.2 Motivação

Tendo em vista o programa de Pós-Graduação em Modelagem Computacional, que é um programa transdisciplinar, assim como a Física Médica, que busca por soluções para diversas áreas relacionadas, esse trabalho tem como motivação e justificção o cerne que acarreta todas as discussões físicas, a análise e a solução de problemas, com a facilitação e automação de processos que visam facilitar e auxiliar o pós-processamento para o diagnóstico médico. Atualmente, a produção de imagens médicas para o diagnóstico de doenças vem enfrentando um problema de *Big Data*, esse problema se desenvolve pelo excesso de imagens geradas. O problema não se dá apenas na produção, mas também no processamento das imagens para os laudos médicos.

Eventualmente, o mercado de profissionais para essa área não supera a produção excessiva, o que necessita de novas formas de auxílio para o manuseio dos bancos de dados e dos processos de avaliação dos processos já realizados.

A automatização de processos computacionais tem desempenhado um papel fundamental no auxílio da organização dos bancos de dados a serem laudados. Essa automatização acarreta vários benefícios aos profissionais, aumentando a produtividade e agilizando os processos de tomadas de decisões quanto às informações já analisadas e armazenadas para análise. As doenças infecciosas respiratórias são umas das mais comuns e acometidas na população brasileira. Junto ao excesso de produção de diagnóstico médico, um processamento dos casos a serem priorizados é necessário para o auxílio mais eficaz e ágil para a administração de tratamentos e gestão do tempo de qualidade dos profissionais.

As redes neurais têm sido utilizadas para auxiliar na gestão e no aprendizado na classificação de padrões das patologias analisadas. O aprendizado dos casos já documentados armazenados pelos profissionais pode auxiliar na separação dos casos a serem laudados. A prévia classificação das imagens com possíveis casos positivos pode agilizar o processo de tratamento e a permanência do tempo na análise de casos de falsos positivos.

As doenças infecciosas inflamatórias acometem boa parte da população brasileira. Entre os tipos de infecções que acometem o trato respiratório inferior, destacam-se as infecções respiratórias agudas (IRA) que, embora se manifestem de formas benignas, são mais importantes

em alguns grupos etários, como crianças e idosos, particularmente são vulneráveis às suas complicações. O diagnóstico tardio dessas patologias pode agravar ainda mais o quadro de recuperação desses pacientes.

A administração do devido tratamento só pode ser administrada após a investigação médica e os laudos clínicos. Portanto, a importância da classificação das imagens no auxílio ao diagnóstico se faz necessário para agilizar os devidos tratamentos.

1.3 Objetivos

1.3.1 Objetivo geral

O objetivo deste trabalho é desenvolver uma solução de aprendizado de máquina para classificar imagens pulmonares e auxiliar no diagnóstico de pneumonia, onde pretende-se **Formalizar a separação dos tipos de pneumonia, desenvolver e treinar um modelo de classificação, utilizando as técnicas de redes neurais.**

1.3.2 Objetivos específicos

Os objetivos específicos se resumem em:

- Escolher e avaliar um modelo para a classificação de imagens pulmonares.
- Avaliar uma solução que possa lidar com grandes volumes de dados de imagens pulmonares para diagnóstico médico.
- Utilizar algoritmos de *machine learning* para adaptar o modelo de classificação para uma diferenciação multiclasse dos tipos de pneumonia.
- Treinar a rede neural para classificar imagens pulmonares em multicategorias (bacteriana, viral, sem pneumonia).

1.4 Estrutura do trabalho

Este trabalho está dividido em 5 capítulos, como segue:

1. No Capítulo 1 está a presente introdução do trabalho. Nela estão contidos a motivação, os objetivos e esta estrutura de organização.
2. O embasamento teórico sobre o tema de pesquisa é apresentado no Capítulo 2. Mais precisamente, são explorados os métodos de Redes neurais e modelagem de problemas de *BigData*, com o intuito de analisar os processos envolvidos no desenvolvimento desse problema de gestão e armazenamento de imagens medicas. Além disso, uma revisão sistemática da literatura é apresentada como método, para determinar e analisar os principais trabalhos que utilizam redes neurais para classificação de imagens medicas.

3. A metodologia do trabalho é apresentada no Capítulo 3. Em um primeiro momento, é apresentada a região de estudo do trabalho, bem como sua importância e principais características. A partir disso, são discutidas detalhadamente as técnicas escolhidas para a modelagem, bem como a organização e sequência das etapas envolvidas no desenvolvimento do trabalho.
4. No Capítulo 4, resultados são apresentados e discutidos. Esses resultados são referentes ao treinamento e a classificação das imagens do *Dataset* onde tem-se as primeiras respostas da validação do modelo escolhido. E ainda, alguns processos do desenvolvimento e adaptação para a classificação multiclasse.
5. As conclusões e considerações finais são apresentadas no Capítulo 5, onde é realizado o fechamento das análises e os próximos trabalho.

2 EMBASAMENTO TEÓRICO

Na introdução, a motivação da pesquisa, bem como os objetivos deste trabalho foram apresentados. Para resolver o problema de pesquisa, o embasamento teórico servirá como fonte de referências, experiências e análises sobre os conceitos de diagnóstico médico, a patologia Pneumonia e o uso inteligência artificial para auxílio de diagnóstico médico e a classificação de padrões de imagens.

2.1 Diagnóstico médico

Segundo DÍAZ NOVÁS; GALLEGO MACHADO; LEÓN GONZÁLEZ (2006) o diagnóstico médico é o processo de identificar e classificar a condição de saúde de um paciente, com base em sinais, sintomas e exames, para determinar o melhor tratamento e prognóstico. O diagnóstico médico, além de ser uma etapa essencial no cuidado ao paciente, envolve a aplicação criteriosa de conhecimentos clínicos, habilidades técnicas e o uso de dados confiáveis.

De acordo com JAMESON et al. (2020), um diagnóstico eficaz requer uma análise cuidadosa de dados precisos e confiáveis. O raciocínio clínico é válido apenas quando baseado em observações precisas; quando esses princípios não são respeitados, os resultados são inevitavelmente equivocados. A validade de qualquer diagnóstico depende da qualidade das informações utilizadas, como dados clínicos, exames laboratoriais e a comunicação entre médico e paciente.

Com base nesses princípios, é importante destacar alguns elementos-chave aplicados no processo de diagnóstico médico:

- Relação médico-paciente: uma relação de confiança e respeito entre médico e paciente é crucial. Esse relacionamento não apenas facilita a obtenção de informações precisas, mas também melhora a cooperação do paciente durante o exame físico e no seguimento das recomendações médicas (RAKEL, 2022).
- Anamnese: A anamnese é a etapa inicial e essencial do processo diagnóstico. Estudos mostram que entre 50 a 75% dos diagnósticos podem ser feitos apenas por meio de uma anamnese detalhada (GOLDMAN; AUSIELLO, 2016). A entrevista médica permite ao médico identificar os sintomas relatados pelo paciente, bem como seu histórico médico, familiar e social, informações fundamentais para guiar o diagnóstico.
- Exame físico: após a anamnese, o exame físico complementa o processo de coleta de informações. Os sinais físicos são manifestações objetivas de condições patológicas e, quando interpretados corretamente, ajudam a confirmar ou refinar hipóteses diagnósticas sugeridas pela anamnese. Sua importância no diagnóstico foi comprovada em diversos estudos clínicos (JAMESON et al., 2020).
- Associação de sintomas e sinais: A correta associação entre sintomas subjetivos e sinais objetivos é um pilar no diagnóstico preciso. O médico precisa correlacionar as queixas

do paciente com as evidências físicas, considerando possíveis padrões de doença, fatores predisponentes e complicações associadas (KUMAR; ABBAS; ASTER, 2020).

- Exames laboratoriais e de imagem: os exames complementares, como exames laboratoriais e de imagem, são fundamentais para confirmar diagnósticos iniciais ou descartar condições suspeitas. O uso adequado dessas ferramentas depende da experiência clínica e do contexto de cada caso (RAHAT, 2021).

A relação médico-paciente é fundamental para o processo de diagnóstico. Uma boa relação médico-paciente proporciona uma visão mais clara dos sintomas relatados, garantindo um ambiente de confiança que facilita tanto o exame físico quanto a cooperação do paciente. Além disso, a anamnese desempenha um papel fundamental no diagnóstico, pois é por meio dela que se obtém grande parte das informações necessárias para o diagnóstico. Estudos mostram que a maioria dos diagnósticos pode ser determinada por uma anamnese bem conduzida (GOLDMAN; AUSIELLO, 2016).

Os sinais físicos são evidências objetivas e verificáveis das doenças. Esses sinais, muitas vezes, confirmam alterações funcionais ou estruturais sugeridas pela anamnese. A combinação desses achados, quando cuidadosamente interpretada, é essencial para a construção de uma hipótese diagnóstica precisa.

Devido à complexidade do corpo humano e à variedade de patologias, o diagnóstico de certas doenças requer a participação de especialistas em suas áreas específicas. A escolha de técnicas diagnósticas específicas, como exames laboratoriais e de imagem, depende do tipo de doença suspeita e dos mecanismos envolvidos nos sintomas apresentados. A utilização eficiente dessas ferramentas garante uma maior precisão no diagnóstico e possibilita tratamentos mais adequados e personalizados para os pacientes.

Em resumo, o diagnóstico médico envolve uma combinação de ciência, arte e interação humana. O processo envolve a aplicação de conhecimentos clínicos, habilidades técnicas e a utilização de dados confiáveis para determinar a condição de saúde de um paciente. A relação médico-paciente é essencial para o processo de diagnóstico, pois permite a obtenção de informações precisas e a construção de uma hipótese diagnóstica precisa. Além disso, a utilização de técnicas diagnósticas específicas e a interpretação cuidadosa dos resultados são essenciais para garantir uma maior precisão no diagnóstico e possibilitar tratamentos mais adequados e personalizados para os pacientes.

2.2 Pneumonia

As doenças respiratórias afetam indivíduos de todas as idades, sendo uma das principais causas de mortalidade global. De acordo com a Organização Mundial da Saúde (HANIFI; MAHMOOD; BHUIYA, 2014), essas doenças representam cerca de 14% das mortes em todo o mundo. Entre elas, as infecções do trato respiratório inferior destacam-se, com taxas que variam

de 31 mortes por 100 mil habitantes em países de alta renda a 91 mortes por 100 mil habitantes em nações de baixa renda (FERRAZ; OLIVEIRA-FRIESTINO; FRANCISCO, 2017).

Dentre as infecções do trato respiratório inferior, as infecções respiratórias agudas (IRA) são particularmente relevantes. Embora geralmente apresentem-se de forma benigna, elas podem levar a complicações graves em grupos etários vulneráveis, como crianças e idosos.

Segundo FERRAZ; OLIVEIRA-FRIESTINO; FRANCISCO (2017), as IRA estão entre as principais causas de mortalidade infantil. Segundo a OMS, em 2013, aproximadamente 3,257 milhões de crianças menores de cinco anos faleceram em decorrência de doenças infecciosas, sendo 14,9% dessas mortes causadas por pneumonia. A pneumonia adquirida na comunidade (PAC) está relacionada a vários fatores de risco, incluindo envelhecimento, tabagismo, doença pulmonar obstrutiva crônica (DPOC), insuficiência cardíaca, colonização da orofaringe, aspiração de conteúdo gástrico, alcoolismo, cirrose hepática, desnutrição, imunossupressão e fatores ambientais. O impacto social da pneumonia é significativo, como demonstram os dados de morbidade e mortalidade.

A pneumonia é um processo inflamatório agudo ou crônico do parênquima pulmonar, causado por bactérias, vírus, fungos ou outros agentes que resultam na inflamação ou infecção do aparelho respiratório. A palavra pneumonia tem origem no grego, e refere-se à inflamação pulmonar, muitas vezes associada à infecção por *Streptococcus pneumoniae* (pneumococo). Os primeiros relatos de pneumonia datam de 1250 a.C., em múmias encontradas no Egito, e desde então a doença tem se disseminado globalmente (MATOSO; CASTRO, 2013).

Segundo ASSUNÇÃO; PEREIRA; ABREU (2018), os avanços nos métodos de diagnóstico têm contribuído para a redução expressiva das mortes por pneumonia nos últimos anos. Contudo, a América Latina ainda concentra países com altos índices de casos da doença, com o Brasil ocupando a primeira posição no ranking. Pacientes internados em Unidades de Terapia Intensiva (UTI), principalmente idosos e crianças, estão entre os principais grupos de risco. A prescrição e o uso adequado de antibióticos são essenciais para evitar o aumento da multirresistência, um desafio crescente na saúde pública.

Embora a maioria dos pacientes com pneumonia recupere-se sem complicações, alguns evoluem para quadros graves, com risco de óbito. A evolução clínica varia conforme a virulência dos agentes etiológicos e o estado imunológico do paciente. O uso de antibióticos para infecções bacterianas e de antivirais, quando indicados, pode acelerar a recuperação, ao reduzir a carga patogênica e modular a resposta imunológica. Neutrófilos, linfócitos e monócitos circulantes desempenham papel crucial na patogênese da pneumonia, e alterações nesses parâmetros podem refletir a gravidade das lesões pulmonares.

A patogênese da pneumonia varia conforme o agente etiológico. Pacientes com pneumonia bacteriana típica, por exemplo, apresentam sintomas mais graves, como leucocitose, neutrofilia com bastonados e bacteremia. Nas fases iniciais da pneumonia, os neutrófilos ativados e os fagócitos mononucleares predominam. Esses mediadores, incluindo enzimas proteolíticas e citocinas, estão associados às lesões pulmonares do hospedeiro (LEE et al., 2010).

Em casos de pneumonia por *Mycoplasma pneumoniae* (MP) e em pneumonias virais as-

sociadas ao sarampo, síndrome respiratória aguda grave (SRAG) ou influenza, a leucopenia e linfopenia são características comuns. A infiltração de células imunes, como linfócitos T, é significativa nas lesões pulmonares iniciais, e as citocinas pró-inflamatórias dessas células podem contribuir para o dano pulmonar (LEE, 2008).

Em conclusão, as infecções respiratórias, incluindo a pneumonia, continuam a ser um dos maiores desafios de saúde pública, afetando de forma desproporcional crianças, idosos e pacientes com comorbidades. Apesar da redução na mortalidade em alguns países, especialmente devido ao avanço no diagnóstico e ao uso adequado de antibióticos, a taxa de mortalidade permanece elevada em nações de baixa renda (ASSUNÇÃO; PEREIRA; ABREU, 2018).

A prevenção, por meio da vacinação, melhorias nas condições de higiene e controle de fatores de risco, como tabagismo e desnutrição, permanece a estratégia mais eficaz para reduzir a morbidade e mortalidade (FERRAZ; OLIVEIRA-FRIESTINO; FRANCISCO, 2017). Além disso, o uso racional de antibióticos é fundamental para evitar o surgimento de cepas multirresistentes, uma ameaça crescente à saúde global.

O progresso na pesquisa de novos métodos terapêuticos e no entendimento da patogênese dessas infecções possibilitará abordagens mais eficazes e personalizadas no tratamento da pneumonia. Contudo, o fortalecimento das políticas de saúde pública e o acesso equitativo a tratamentos e à prevenção em escala global são essenciais para mitigar o impacto dessas doenças nos próximos anos.

2.3 Laudo médico

O laudo médico é um documento técnico elaborado por um profissional de saúde, no qual são descritos os achados clínicos, exames realizados e a conclusão diagnóstica sobre a condição de um paciente. Ele tem como objetivo fornecer informações detalhadas sobre o estado de saúde do indivíduo, auxiliando no diagnóstico, tratamento e acompanhamento médico. Além disso, o laudo pode ser utilizado para fins jurídicos, periciais e administrativos, servindo como prova documental em processos legais, atestados médicos ou solicitações de benefícios. Sua elaboração deve ser clara, objetiva e fundamentada em evidências clínicas e exames complementares, garantindo precisão e confiabilidade nas informações prestadas (ARRUDA et al., 2000).

As imagens radiográficas de pneumonia apresentam características distintas conforme sua etiologia, sendo fundamentais para o diagnóstico diferencial entre infecções bacterianas e virais. Na pneumonia bacteriana conforme Figura 1, observa-se uma consolidação lobar densa e bem delimitada, frequentemente acompanhada de broncograma aéreo, evidenciando a substituição do ar alveolar por exsudato inflamatório, além da possibilidade de derrame pleural em casos mais avançados (FELSON; GOODMAN (2014); SIMON (1997)).

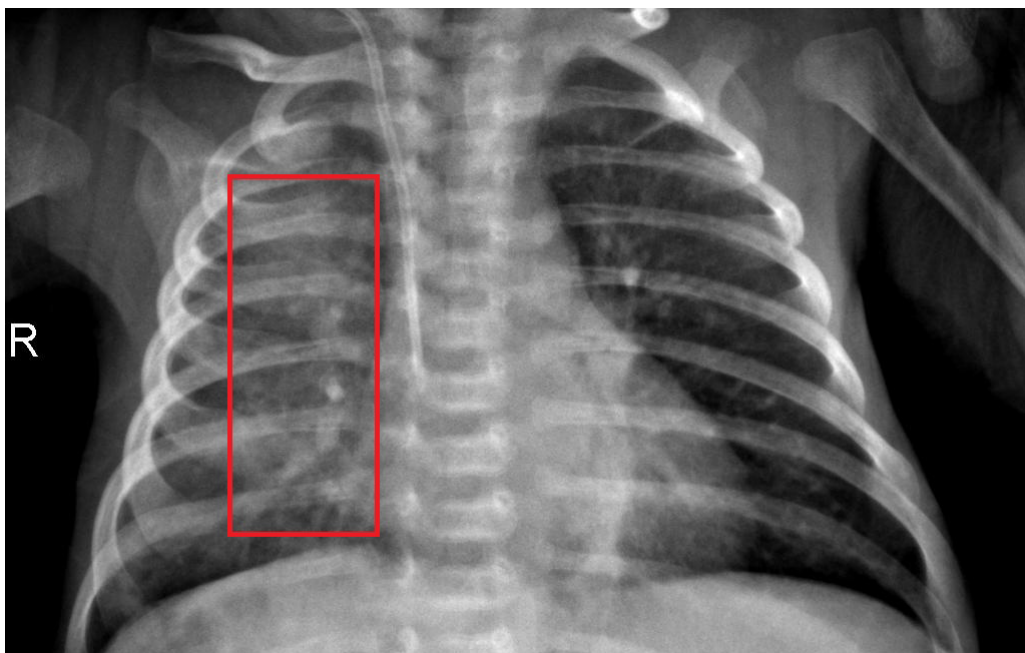


Figura 1: Imagem de pneumonia bacteriana.

Já a pneumonia viral conforme Figura 2 se manifesta predominantemente com um padrão intersticial difuso, caracterizado por opacidades em vidro fosco e acometimento bilateral, refletindo a inflamação e o espessamento dos septos interalveolares, sem a formação de consolidações extensas MURRAY; NADEL (2021); KAUCZOR; KANNE; LYNCH (2017).

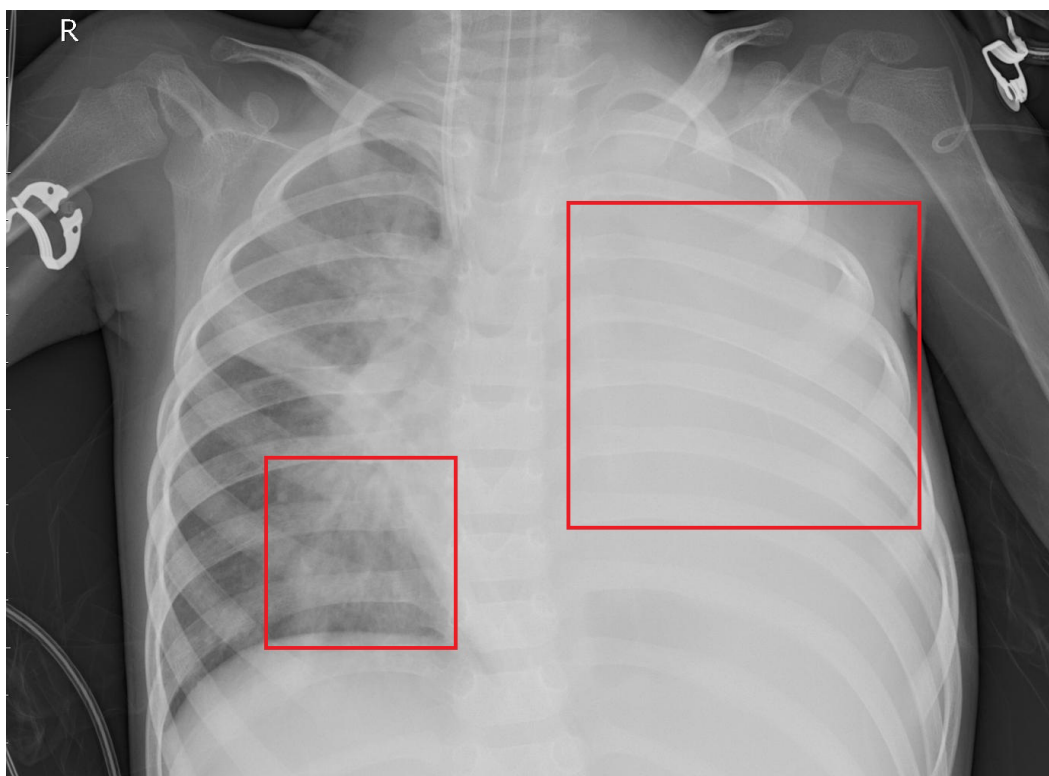


Figura 2: Imagem de pneumonia viral.

Essas distinções radiológicas são essenciais para a correta avaliação do quadro clínico, au-

xiliando na escolha da conduta terapêutica mais adequada (TISIOLOGIA, 2021).

2.4 Inteligencia artificial no diagnóstico médico

O artigo de ARAÚJO; HORNUNG (2022) aborda o uso da inteligência artificial (IA) no diagnóstico de doenças e as implicações jurídicas relacionadas à responsabilidade civil do médico em casos de erro diagnóstico. É discutido como a IA tem sido utilizada para auxiliar profissionais de saúde na tomada de decisões clínicas, destacando suas vantagens, como maior precisão e eficiência. No entanto, analisam também os desafios legais decorrentes do uso dessas tecnologias, especialmente no que se refere à atribuição de responsabilidade quando ocorre um equívoco no diagnóstico. O estudo explora se a culpa recai exclusivamente sobre o médico, que utilizou a IA como ferramenta de apoio, ou se há a possibilidade de responsabilização dos desenvolvedores e fornecedores da tecnologia. Assim, a pesquisa contribui para o debate sobre a regulamentação da IA na área médica e a necessidade de diretrizes que garantam a segurança dos pacientes sem desonerar completamente os profissionais da saúde.

2.5 Redes neurais artificiais

A estrutura do cérebro humano, particularmente sua rede de neurônios interconectados, inspirou a criação das Redes Neurais Artificiais (RNA). Assim como o cérebro processa informações complexas e se adapta a novas experiências, as RNAs são projetadas para aprender e adaptar-se por meio do processamento de dados.

O conceito de RNA foi introduzido por MCCULLOCH; PITTS (1943), que desenvolveram um modelo computacional simplificado inspirado nos neurônios biológicos. Eles demonstraram como cálculos complexos podem ser realizados por meio da lógica proposicional, utilizando neurônios artificiais. Esse trabalho pioneiro foi um marco inicial no campo da inteligência artificial, abrindo caminho para o desenvolvimento de novos algoritmos e técnicas para resolver problemas complexos (GERON, 2021).

Antes de discutir as redes neurais artificiais em detalhes, é útil compreender a estrutura de um neurônio biológico. Como ilustrado na Figura 3, um neurônio biológico possui um corpo celular, dendritos que recebem sinais, e um axônio que transmite impulsos elétricos. Na extremidade do axônio, os terminais sinápticos se conectam a outros neurônios ou células-alvo (MCCULLOCH; PITTS, 1943). Essa estrutura inspirou o conceito de neurônio artificial.

Com base nessa inspiração, MCCULLOCH; PITTS (1943) criaram o neurônio artificial, que opera de maneira semelhante. Esse modelo utiliza entradas ponderadas, calcula a soma dessas entradas e gera uma saída binária quando a soma atinge um limiar específico.

Posteriormente, ROSENBLATT (1958) avançou o campo ao introduzir o Perceptron, uma arquitetura simples de RNA baseada em uma Unidade Lógica de Limiar (TLU). A TLU aplica uma função degrau à soma ponderada das entradas, gerando uma saída binária (Figura 4). Esta foi a primeira aplicação prática das RNAs em problemas de classificação simples.

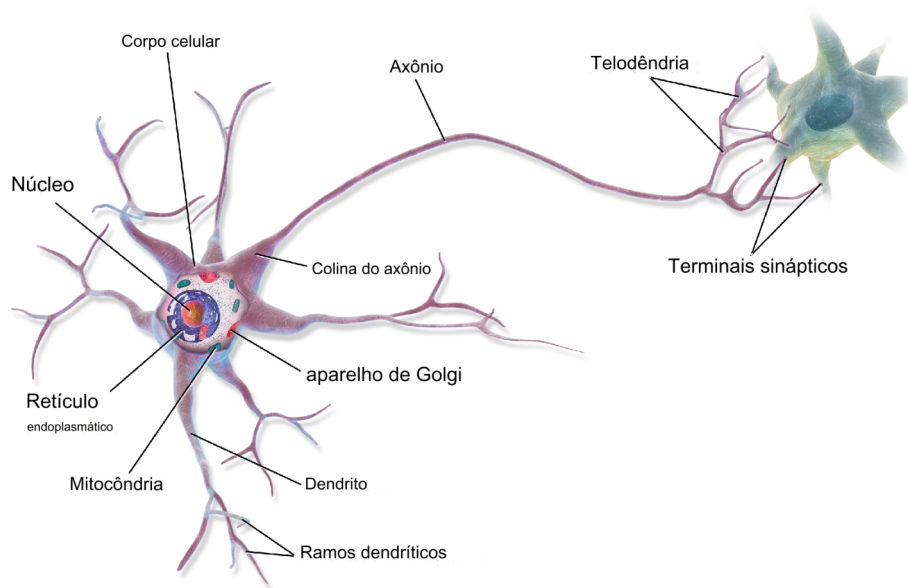


Figura 3: Neurônio Biológico.

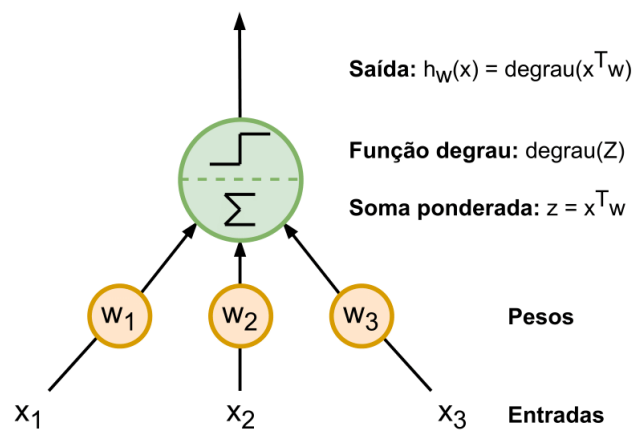


Figura 4: Unidade lógica de Limiar (TLU): Neurônio Artificial.

No entanto, as limitações do Perceptron logo se tornaram evidentes. MINSKY; PAPERT (1969) mostraram que ele não era capaz de resolver problemas como o XOR (Exclusive OR), uma função não linearmente separável, como visto na Figura 5. Para superar essa limitação, surgiram os Perceptrons Multicamadas (MLP), que introduziram camadas ocultas e permitiram resolver problemas complexos como o XOR.

As MLPs, que possuem múltiplas camadas de Unidades Lógicas de Limiar, são classificadas como Redes Neurais Profundas (DNN) quando contêm várias camadas ocultas. As DNNs têm a capacidade de aprender representações hierárquicas de dados, o que as torna adequadas para resolver problemas avançados de classificação e reconhecimento de padrões (HINTON; OSINDERO; TEH, 2006).

Com o avanço das DNNs, surgiu o campo do *Deep Learning*, que revolucionou áreas como o processamento de imagens e a análise de grandes volumes de dados. Um dos principais exemplos dessa evolução é a Convolutional Neural Network (CNN), uma arquitetura projetada para

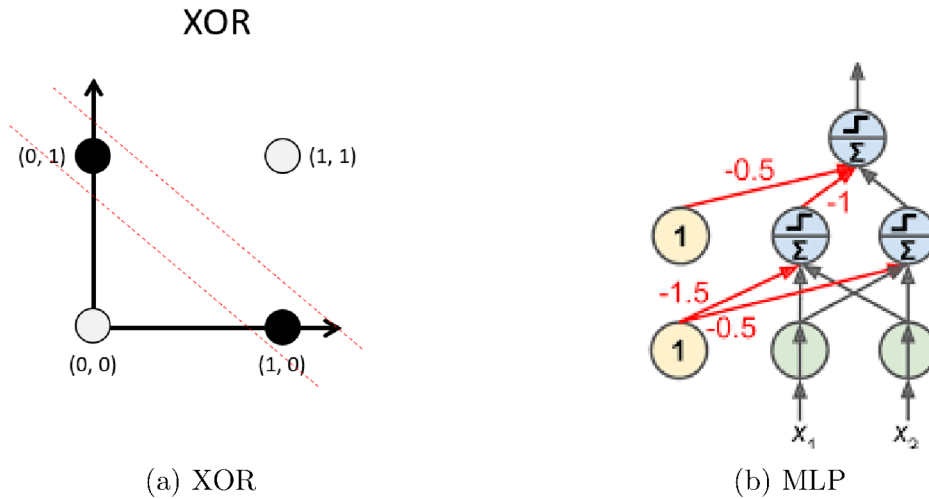


Figura 5: Problema de classificação XOR e uma MLP que soluciona.

o reconhecimento de padrões em imagens Figura 6. As CNNs utilizam camadas convolucionais para detectar características locais nas imagens, permitindo o reconhecimento de objetos complexos.

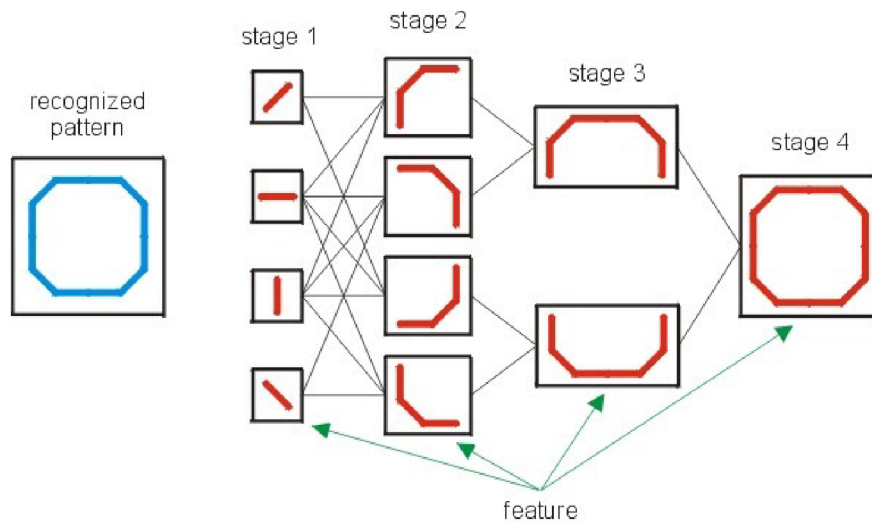


Figura 6: Rede Neural Profunda (CNN) aplicando convolução para reconhecer padrões em imagens.

Hoje, as RNAs são amplamente utilizadas em aplicações como reconhecimento de voz, previsão de séries temporais e processamento de linguagem natural. No entanto, ainda enfrentam desafios, como a necessidade de grandes conjuntos de dados para treinamento e o ajuste fino dos hiperparâmetros para otimizar o desempenho do modelo (ZHONG et al., 2016).

As Redes Neurais Artificiais evoluíram significativamente desde suas primeiras concepções. Atualmente, elas desempenham um papel essencial em diversas áreas, fornecendo soluções eficazes para problemas complexos. Com a contínua evolução da pesquisa, novas arquiteturas e técnicas surgem, aprimorando ainda mais a capacidade das RNAs de lidar com desafios de maior complexidade.

2.6 Redes neurais de classificação binárias

A classificação de padrões em redes neurais binárias tratam a modelagem do problema como uma forma de simplificar como os aprendizado dos números binários 0 e 1. O objetivo da classificação binária é aprender uma função $F(x)$ que minimize a probabilidade de classificação incorreta $PyF(x) < 0$, onde y é o rótulo da classe com 1 para positivo e -1 para negativo. Existem muitos métodos influentes de classificação binária, como métodos *kernel* HOFMANN; SCHÖLKOPF; SMOLA (2008), métodos ensemble POLIKAR (2006) e métodos de aprendizado profundo BENGIO (2009). A máquina de vetores de suporte (SVM) VAPNIK (1999) é um método *kernel* clássico. Os métodos ensemble incluem *boosting* FREUND; SCHAPIRE (1997); FRIEDMAN; HASTIE; TIBSHIRANI (2000) e *random forest* (RF) BREIMAN (2001). Os métodos de aprendizado profundo são baseados em redes neurais artificiais (ANNs) BISHOP (1995).

2.7 Redes neurais de classificação multiclasse

Os classificadores multiclasse também conhecidos como classificadores multinominais, diferente dos binários podem distinguir entre mais de duas classes. Alguns algoritmos de classificadores como (RF) BREIMAN (2001) e *Naive Bayes* AQUINO; FONSECA; OLIVEIRA (2016) conseguem lidar com diversas classes, outros classificadores como de regressão logística ou máquina de vetores LAFETÁ et al. (2020) são estritamente classificadores binários.

2.8 Aspectos Matemáticos e Computacionais das Redes Neurais Convolucionais (CNNs)

As Redes Neurais Convolucionais (CNNs) são um tipo de rede neural projetada para lidar com dados em formato de grade, como imagens. Elas se destacam por sua capacidade de aprender padrões espaciais automaticamente, reduzindo a necessidade de extração manual de características. A arquitetura de uma CNN é composta, essencialmente, por três tipos principais de camadas: **camadas convolucionais**, **camadas de pooling** e **camadas totalmente conectadas**.

2.9 Camadas Convolucionais

A principal operação realizada pelas CNNs é a *convolução*, que pode ser descrita matematicamente como:

$$S(i, j) = \sum_m \sum_n X(i + m, j + n) K(m, n) \quad (2.1)$$

onde:

- $X(i, j)$ representa a matriz de entrada (imagem),
- $K(m, n)$ é o filtro (kernel) aplicado à imagem,
- $S(i, j)$ representa a saída do mapa de características.

Cada filtro convolucional aprende um conjunto específico de características, como detecção de bordas, texturas e padrões mais complexos conforme as camadas aumentam.

2.10 Camadas de Pooling

Após a convolução, é comum aplicar uma camada de *pooling* para reduzir a dimensionalidade dos mapas de características. O **Max Pooling**, um dos métodos mais utilizados, é definido por:

$$Y(i, j) = \max_{m, n} X(2i + m, 2j + n) \quad (2.2)$$

Essa operação reduz a sensibilidade a deslocamentos pequenos e melhora a eficiência computacional.

2.11 Camadas Totalmente Conectadas

Após a extração de características, os mapas de ativação são achatados e passados para camadas *fully connected*. Essas camadas aplicam transformações lineares seguidas por funções de ativação, como a ReLU ou Softmax para classificação:

$$z = Wx + b \quad (2.3)$$

$$\hat{y} = \frac{e^{z_i}}{\sum_j e^{z_j}} \quad (2.4)$$

onde:

- W representa a matriz de pesos da camada,
- x é a entrada achatada do mapa de características,
- b é o viés (bias) é um parâmetro adicional introduzido em cada neurônio da rede para aumentar sua flexibilidade e permitir um melhor ajuste da função de ativação,
- $\hat{y}$ representa as probabilidades de cada classe.

2.12 Treinamento da Rede Neural Convolucional

O treinamento de uma Rede Neural Convolucional (CNN) consiste em ajustar os pesos W e os vieses b da rede para minimizar o erro entre as previsões do modelo e os valores reais dos dados. Esse processo ocorre por meio de três etapas principais: *forward pass*, **cálculo da função de perda** e *backpropagation* com otimização dos parâmetros.

2.12.1 Forward Pass (Propagação para Frente)

Na etapa de forward pass, a entrada x (uma imagem, por exemplo) passa por todas as camadas da rede, onde são aplicadas operações como convolução, pooling e camadas totalmente conectadas. Cada camada transforma os dados de entrada de acordo com seus filtros e pesos aprendidos.

Matematicamente, para cada camada:

$$z = Wx + b$$

onde z representa a saída linear da camada. Em seguida, é aplicada uma função de ativação f , como ReLU:

$$a = f(z)$$

No final da rede, as previsões $\hat{y}$ são obtidas. Em um problema de classificação, isso pode ser feito com a função *softmax*:

$$\hat{y}_i = \frac{e^{z_i}}{\sum_j e^{z_j}}$$

onde $\hat{y}_i$ representa a probabilidade da imagem pertencer à classe i .

2.12.2 Cálculo da Função de Perda

A função de perda mede o erro entre as previsões $\hat{y}$ e os rótulos reais y . Para classificação, a função mais comum é a **entropia cruzada**:

$$\mathcal{L} = - \sum_i y_i \log(\hat{y}_i)$$

onde: - y_i é 1 se a amostra pertence à classe i e 0 caso contrário, - $\hat{y}_i$ é a probabilidade prevista pelo modelo para a classe i .

O objetivo do treinamento é minimizar essa função de perda, tornando as previsões mais próximas dos valores reais.

2.12.3 Backpropagation e Atualização dos Pesos

A etapa de *backpropagation* ajusta os pesos da rede para minimizar a perda. Isso ocorre por meio do **cálculo do gradiente** da função de perda em relação aos pesos W e vieses b .

A atualização dos pesos segue a regra do **gradiente descendente estocástico (SGD)** ou variantes como Adam:

$$W = W - \eta \frac{\partial \mathcal{L}}{\partial W}$$

$$b = b - \eta \frac{\partial \mathcal{L}}{\partial b}$$

onde η é a **taxa de aprendizado**, que controla o tamanho do ajuste em cada iteração.

Esse processo de *forward pass* $\rightarrow$ cálculo da perda $\rightarrow$ *backpropagation* $\rightarrow$ atualização dos pesos é repetido por múltiplas **épocas**, refinando progressivamente a capacidade da CNN de fazer previsões corretas.

2.13 Otimizações no Treinamento

Para melhorar o desempenho do treinamento, algumas técnicas são amplamente utilizadas:

- **Batch Normalization:** Normaliza as ativações dentro da rede para evitar variações extremas nos gradientes.
- **Dropout:** Desativa aleatoriamente neurônios durante o treinamento para evitar *overfitting*.
- **Aumento de Dados (Data Augmentation):** Gera variações artificiais das imagens para tornar a rede mais robusta.
- **Taxa de Aprendizado Adaptativa:** Técnicas como Adam ajustam automaticamente a taxa de aprendizado.

O treinamento de redes neurais convolucionais é um processo complexo que envolve otimização matemática, ajuste de hiperparâmetros e aprendizado estatístico GOODFELLOW; BENGIO; COURVILLE (2016), AGGARWAL (2018), SHANMUGAMANI (2018). O uso de abordagens matemáticas rigorosas, como a propagação para frente e retropropagação, permite que os modelos aprendam padrões complexos a partir dos dados.

A formulação matemática das CNNs desempenha um papel central na eficiência e interpretabilidade desses modelos, permitindo a extração de características relevantes a partir de dados estruturados, como imagens. Operações fundamentais, como convolução, *pooling* e funções de ativação, são descritas por transformações lineares e não lineares que possibilitam a modelagem de padrões complexos. No entanto, a implementação manual dessas operações pode ser computacionalmente onerosa e suscetível a imprecisões numéricas. Para mitigar tais desafios,

o uso de bibliotecas especializadas em *Python*, como *TensorFlow*, *Keras* e *PyTorch*, torna-se essencial, pois essas ferramentas fornecem implementações otimizadas para operações matriciais, diferenciação automática e otimização de funções de perda. Além disso, tais bibliotecas são projetadas para aproveitar arquiteturas de hardware modernas, como GPUs e TPUs, acelerando significativamente o treinamento de modelos profundos. Dessa forma, a combinação entre formalismo matemático rigoroso e infraestrutura computacional eficiente permite avanços na aplicação de CNNs em diversas áreas científicas e tecnológicas, incluindo análise de imagens médicas, reconhecimento de padrões e aprendizado de máquina em larga escala.

2.14 Linguagem e pacotes

Python é uma linguagem de programação de alto nível, interpretada e de propósito geral, criada por Guido van Rossum em 1991. Sua sintaxe é projetada para ser clara e legível, facilitando a escrita e manutenção de código. Além disso, *Python* é uma linguagem interpretada, o que significa que o código-fonte é executado linha por linha pelo interpretador. Isso torna o desenvolvimento e teste de código mais rápidos, embora resulte em uma execução ligeiramente mais lenta em comparação com linguagens compiladas.

A biblioteca padrão do *Python* é extensa e abrangente, facilitando o desenvolvimento de aplicações completas sem a necessidade de importar bibliotecas externas. Além disso, *Python* suporta programação orientada a objetos, programação imperativa e programação funcional, oferecendo flexibilidade aos desenvolvedores para escolher o paradigma que melhor se adapte às suas necessidades (VAN ROSSUM et al., 2007).

Python é amplamente utilizado em campos como ciência de dados e inteligência artificial, devido à disponibilidade de bibliotecas poderosas como *NumPy*, *pandas*, *TensorFlow* e *PyTorch*. Além disso, *Python* é dinamicamente tipada, o que significa que as variáveis não precisam ser explicitamente declaradas com um tipo. Isso oferece flexibilidade, mas também exige atenção ao lidar com tipos de dados (VAN ROSSUM et al., 2007).

O *Keras* é uma biblioteca de código aberto escrita em *Python*, projetada para facilitar o desenvolvimento e a experimentação de redes neurais. Ela fornece uma interface de alto nível para construir, treinar e avaliar modelos de aprendizado profundo, tornando-se uma ferramenta poderosa para os desenvolvedores de inteligência artificial. Algumas características-chave do *Keras* incluem a simplicidade por sua API simples e intuitiva, permitindo aos desenvolvedores criar modelos de forma rápida e eficiente.

O *Keras* também é modular e portátil, suportando diversas arquiteturas de rede neural, incluindo Redes Neurais Convolucionais (CNNs), recorrentes (RNNs) e outros tipos. Isso facilita a implementação de modelos para tarefas específicas, como visão computacional, processamento de linguagem natural, entre outras (GULLI; PAL, 2017).

O treinamento e a avaliação são fundamentais para o desenvolvimento de modelos de aprendizado profundo, e o *Keras* fornece métodos simples para treinar modelos com dados de treinamento e avaliá-los com dados de teste. Além disso, oferece funcionalidades para monitoramento

do desempenho durante o treinamento e integração com o *TensorFlow* (GULLI; PAL, 2017).

Embora seja independente, o Keras está profundamente integrado ao *TensorFlow*, sendo inclusive a API de alto nível padrão para construção de modelos no *TensorFlow* 2.0 e versões posteriores. Isso torna o Keras uma ferramenta essencial para os desenvolvedores que trabalham com o *TensorFlow*.

O *TensorFlow* é uma biblioteca de código aberto para aprendizado de máquina e aprendizado profundo, desenvolvida pelo Google, que utiliza um modelo de programação gráfico para representar as operações como nós em um grafo e as arestas representam os dados que fluem entre esses nós. Isso permite a otimização de operações complexas e a visualização de fluxos de dados, tornando-o uma ferramenta poderosa para a construção de modelos de aprendizado profundo.

A arquitetura do *TensorFlow* é baseada em um modelo de programação gráfico, que permite a definição de operações como nós em um grafo e a execução dessas operações em paralelo. Isso permite a escalabilidade e a flexibilidade do *TensorFlow*, tornando-o adequado para uma ampla gama de aplicações e tamanhos de conjuntos de dados (PANG; NIJKAMP; WU, 2020).

Além disso, o *TensorFlow* oferece uma variedade de ferramentas e funcionalidades para a construção de modelos de aprendizado profundo, incluindo suporte a redes neurais profundas, otimização de hiper parâmetros e geração automática de arquiteturas de modelo. O *TensorFlow* também é integrado com outras bibliotecas e *frameworks*, como o Keras e o *TensorFlow Lite*, permitindo a construção de modelos mais simples e intuitivos (PANG; NIJKAMP; WU, 2020).

2.15 *Visual Geometry Group (VGG)*

O *Visual Geometry Group* (VGG) é uma rede neural profunda projetada para a tarefa de reconhecimento de imagens, desenvolvida pelo grupo de pesquisa Visual Geometry Group da Universidade de Oxford. A VGG é conhecida por sua arquitetura simples e eficaz, que se tornou um padrão para a construção de redes neurais para tarefas de visão computacional.

A VGG é composta por uma série de camadas convolucionais, seguidas de camadas de pooling e, finalmente, camadas densas. A arquitetura da VGG é dividida em várias versões, incluindo VGG16 e VGG19, que diferem apenas no número de camadas convolucionais.

Visão geral da arquitetura da VGG16:

Camadas convolucionais:

- 2 camadas convolucionais com 64 filtros de tamanho 3x3.
- 2 camadas convolucionais com 128 filtros de tamanho 3x3.
- 3 camadas convolucionais com 256 filtros de tamanho 3x3.
- 3 camadas convolucionais com 512 filtros de tamanho 3x3.
- 3 camadas convolucionais com 512 filtros de tamanho 3x3.

Camadas de pooling:

- 5 camadas de pooling com tamanho de janela 2x2 e stride 2.

Camadas densas:

- 2 camadas densas com 4096 neurônios.
- 1 camada densa com 1000 neurônios (saída).

A VGG é conhecida por suas seguintes características:

- **Arquitetura simples:** A VGG tem uma arquitetura simples e fácil de entender, o que a torna uma boa opção para iniciantes em visão computacional.
- **Eficácia:** A VGG é muito eficaz em tarefas de reconhecimento de imagens, especialmente em datasets como o ImageNet.
- **Flexibilidade:** A VGG pode ser facilmente adaptada para diferentes tarefas de visão computacional, como detecção de objetos e segmentação de imagens.
- **Reprodutibilidade:** A VGG é uma das redes neurais mais reproduzíveis, o que significa que é fácil de reproduzir os resultados obtidos com a rede.

A VGG tem sido amplamente utilizada em várias aplicações, incluindo:

- **Reconhecimento de imagens:** A VGG é uma das redes neurais mais populares para tarefas de reconhecimento de imagens, especialmente em datasets como o ImageNet.
- **Detecção de objetos:** A VGG pode ser utilizada para detecção de objetos em imagens, especialmente quando combinada com outras técnicas de visão computacional.
- **Segmentação de imagens:** A VGG pode ser utilizada para segmentação de imagens, especialmente em aplicações médicas.
- **Geração de imagens:** A VGG pode ser utilizada para gerar imagens sintéticas, especialmente em aplicações de realidade virtual e aumentada.

A VGG é uma rede neural profunda muito popular e eficaz para tarefas de visão computacional, especialmente em reconhecimento de imagens. Sua arquitetura simples e flexibilidade a tornam uma boa opção para iniciantes e especialistas em visão computacional.

Usando como base os modelos de redes neurais para identificação de estruturas conhecidas, a rede segue a ideia de estrutura de neurônios conforme a figura 7. Observa-se que os dados de entrada são processados pelos neurônios localizados na primeira camada oculta. Esses neurônios, por sua vez, geram uma saída que serve como entrada para a camada seguinte, e esse processo continua em cascata, culminando na obtenção da saída final. Cada camada pode conter um ou vários neurônios.

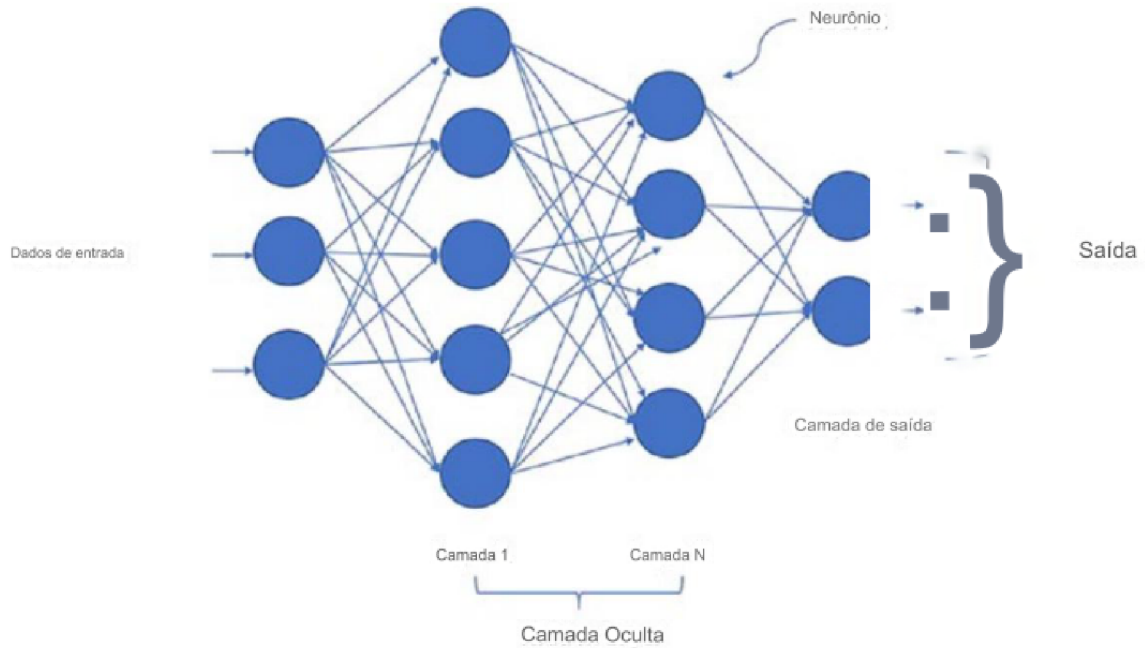


Figura 7: Exemplo de estrutura neural.

O modelo de rede neural denominado de *VGG/VGGNet* como modelo base para o entendimento e funcionamento da rede neural que idealizamos, o modelo VGG suporta 16 camadas, por esse motivo também é conhecido como VGG16. Esse modelo foi desenvolvido por KERMANY; ZHANG; GOLDBAUM (2018) o artigo intitulado "*Large Scale Visual Recognition Challenge* (Redes Convolucionais Muito Profundas para Reconhecimento de Imagem em grande escala)" o modelo é demonstrado na figura 8 SIMONYAN; ZISSERMAN (2014) e no artigo *Pneumonia detection in chest X-ray images using convolutional neural network and transfer learning* por (JAIN et al., 2020).

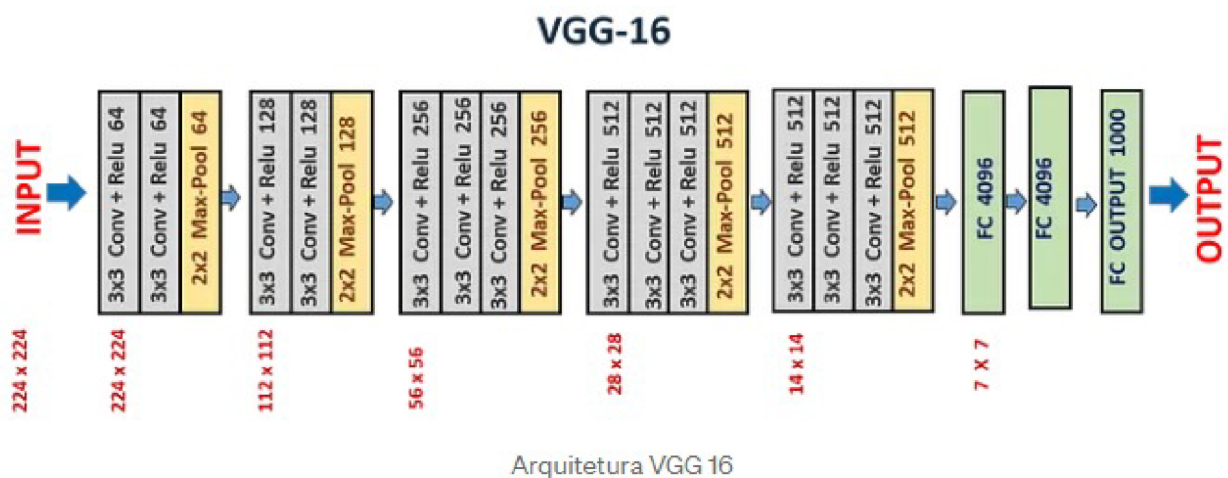
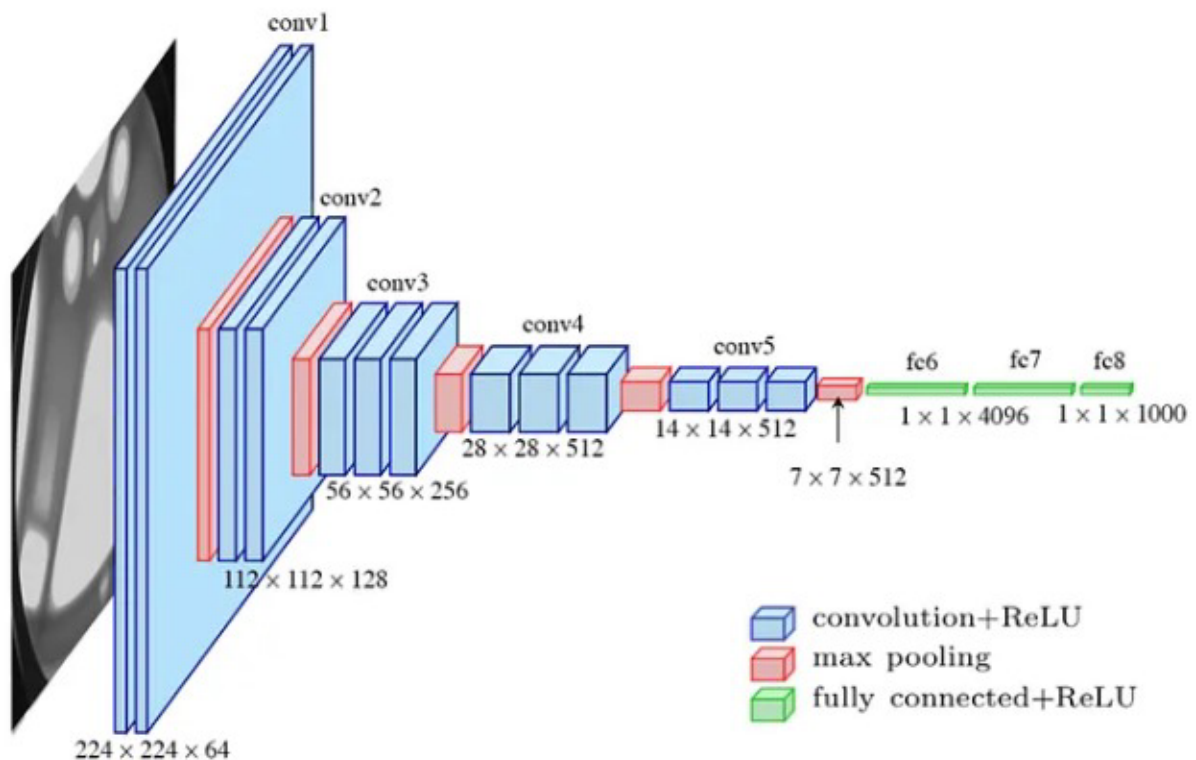


Figura 8: Arquitetura do modelo VGG-16.

O VGG16, treinado no conjunto de dados ImageNet, atingiu uma precisão média de implantação de cerca de 92,7% em cinco testes. ImageNet consiste em mais de 14 milhões de imagens distribuídas em quase 1.000 categorias exclusivas. VGG16 foi um dos modelos proeminentes exibidos no ILSVRC-2014 *ImageNet Large Scale Visual Recognition Challenge 2014*, superando consideravelmente o desempenho do *AlexNet*. Isso foi alcançado substituindo vários filtros 3×3 por filtros maiores. Para treinar o modelo VGG16.

A arquitetura do *VGGNet-16* tem 16 camadas e pode classificar imagens em até 1000 categorias, como imagens de teclado, animais, lápis, mouse, etc. O VGG16 pode processar imagens de 224×224 píxeis, conforme Figura 9 (SIMONYAN; ZISSERMAN, 2014).



Arquitetura VGG-Net

Figura 9: Arquitetura do modelo VGG-16.

As dimensões de altura e largura são diminuídas e redimensionadas por camadas *pooling*, essas camadas entram em ação após uma série de camadas convolucionais, os filtros gerados a cada passada das camadas aumentam em múltiplos de 2, começando com 64 filtros e parando em 512 filtros na última camada, de forma sistemática conforme a figura 9.

O modelo *VGG* possui 5 configurações diferentes Chamadas de A até E conforme a Figura 10, para cada configuração são accionando mais camadas e filtros sucessivamente, cada modo de configuração possui uma quantidade de camadas de

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figura 10: Configurações do modelo VGG-16.

A cada camada, as imagens são reduzidas a dados, os dados são analisados como uma matriz, os pontos da matriz são os pixels da imagem. A redução ocorre convertendo uma matriz 3×3 para uma 2×2 assim por diante, essa redução facilita a análise de padrões conhecidos e aprendidos para análise e comparação com as imagens do banco de dados.

2.16 Trabalhos relacionados

O estudo de TSIKNAKIS et al. (2020), aborda o desenvolvimento de um framework de inteligência artificial (IA) que não apenas realiza a triagem de COVID-19 usando imagens de raios-x torácicos, mas também oferece interpretabilidade dos resultados, o que é crucial para a aceitação clínica e a tomada de decisão. A metodologia empregada para o estudo foi a utilização de técnicas avançadas de aprendizado profundo, particularmente redes neurais convolucionais (CNNs) conforme a Figura 11. Para a análise das imagens, para garantir a interpretabilidade, o framework incorpora métodos de explicação, como mapas de ativação e técnicas baseadas em gradientes, que ajudam a visualizar quais regiões da imagem influenciam a decisão do modelo. O estudo conclui que a combinação de IA com interpretabilidade é essencial para a aceitação clínica. Modelos de IA que oferecem explicações claras podem aumentar a confiança dos médicos nos resultados automatizados e melhorar a integração dessas tecnologias na prática clínica. Os resultados do estudo demonstram que o treinamento resultou em uma eficácia de Acurácia

Binária (100%), Acurácia Ternária (85%) e Quaternária Acurácia (76%). O trabalho destaca a importância de não apenas criar modelos de IA precisos, mas também garantir que eles sejam compreensíveis e transparentes para os usuários finais, promovendo uma maior confiança e adoção dessas tecnologias na medicina.

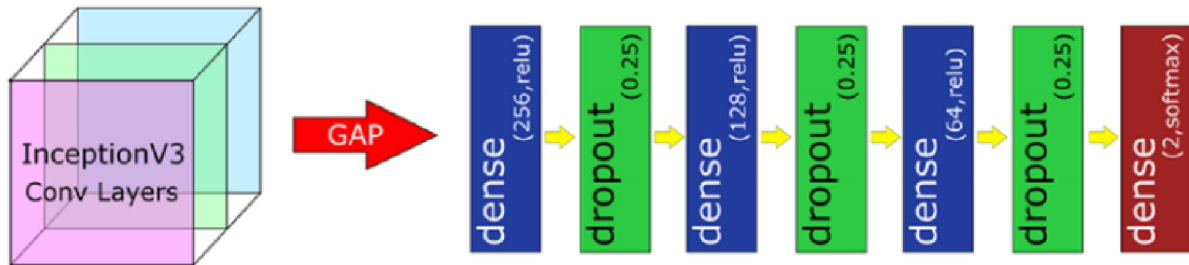


Figura 11: Rede Neural Artificial proposta por TSIKNAKIS et al. (2020).

O estudo de IBRAHIM et al. (2021) explora a aplicação de técnicas de aprendizado profundo para classificar pneumonia em imagens de raios-x torácicos, com um foco particular na pneumonia associada à COVID-19. O estudo utiliza (CNNs), uma arquitetura de aprendizado profundo eficaz para análise de imagens. O modelo é treinado com um *dataset* conforme mostra a Figura 12 que inclui imagens de raios-x de pacientes com pneumonia viral, bacteriana e COVID-19, para permitir uma classificação precisa das diferentes formas de pneumonia. O estudo apresenta métricas de desempenho do modelo, como precisão, sensibilidade, especificidade e outras, demonstrando a eficácia do modelo na classificação correta das imagens de raios-x. O estudo conclui que técnicas de aprendizado profundo são promissoras para a detecção e classificação de pneumonia a partir de imagens de raios-x. Os resultados do estudo apresentam uma eficácia de Acurácia Binária (99.71%) e Acurácia Ternária (95%). A capacidade de distinguir entre pneumonia COVID-19 e outras formas de pneumonia pode melhorar significativamente o diagnóstico e o tratamento, especialmente em contextos de alta demanda durante pandemias. O trabalho enfatiza o potencial dos métodos de aprendizado profundo para auxiliar os profissionais de saúde na triagem e diagnóstico de doenças pulmonares, especialmente durante crises sanitárias como a pandemia de COVID-19.

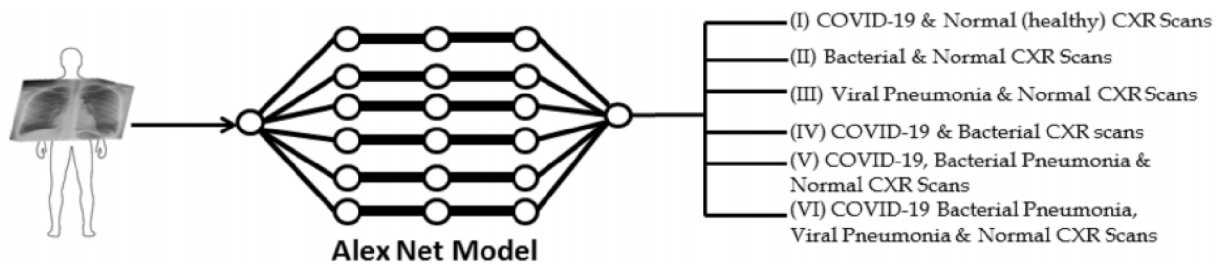


Figura 12: Rede Neural Artificial proposta por IBRAHIM et al. (2021).

O estudo de BHANDARY et al. (2020) investiga o uso de frameworks de aprendizado profundo (deep learning) para detectar anomalias pulmonares em imagens de radiografias de tórax

e tomografias computadorizadas (CT scans) dos pulmões. O trabalho se concentra na criação e avaliação de modelos de aprendizado profundo que podem analisar imagens de raios-x e CT scans para identificar e classificar anomalias pulmonares conforme a Figura 13, como nódulos e outras condições patológicas. O estudo visa melhorar a precisão e a eficiência na detecção precoce de doenças pulmonares, contribuindo para diagnósticos mais rápidos e precisos, utilizando (CNNs), um tipo de rede neural profunda eficaz para tarefas de visão computacional, para analisar e classificar imagens de raios-x e CT scans, os resultados apresentados são de Acurácia Câncer (97.27%) e Pneumonia Acurácia (96,8%). O estudo destaca a eficácia dos frameworks de aprendizado profundo na detecção de anomalias pulmonares e discute o potencial impacto desses modelos na prática clínica, incluindo possíveis melhorias no diagnóstico e tratamento de doenças pulmonares.

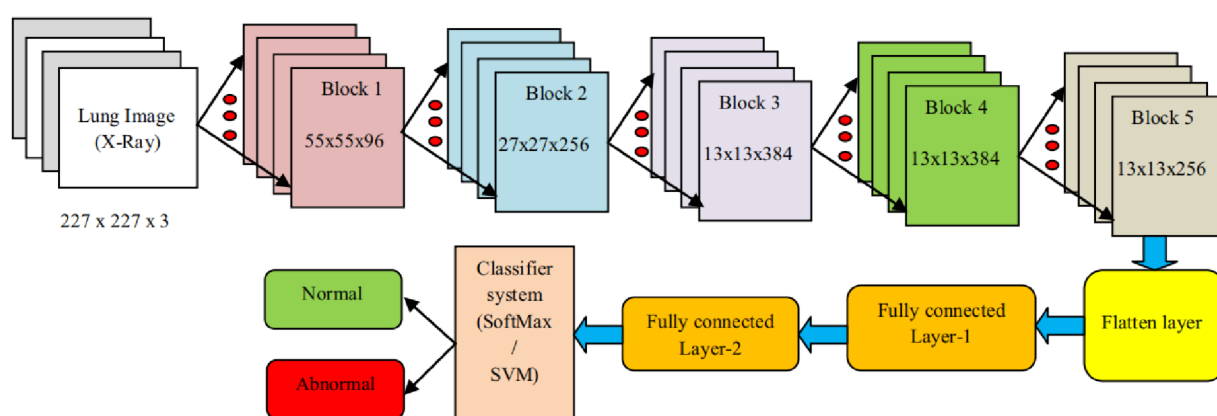


Figura 13: Rede Neural Artificial proposta por BHANDARY et al. (2020).

O estudo de REZAEI; BADIEI; MESHGINI (2020) estuda uma abordagem híbrida de aprendizado profundo e aprendizado por transferência para a classificação de COVID-19 em imagens de raios-x torácicos. O estudo utiliza (CNNs) para análise de imagens e na extração de características relevantes. Para melhorar o desempenho e reduzir o tempo de treinamento, o modelo adota uma abordagem de aprendizado por transferência, utilizando redes pré-treinadas em grandes conjuntos de dados de imagens como IMAGENET conforme a Figura 14. Essas redes são ajustadas com um conjunto específico de dados de raios-x torácicos para o problema de COVID-19. A abordagem híbrida de aprendizado profundo e aprendizado por transferência se mostra eficaz para a detecção de COVID-19 em imagens de raios-x, superando algumas limitações dos métodos tradicionais e de modelos menos sofisticados, o estudo apresenta os resultados de Acurácia Binária (99%) e Acurácia Multi-classe (94%). O estudo sugere que a combinação dessas técnicas pode proporcionar uma ferramenta poderosa para a triagem e diagnóstico rápido de COVID-19, o que é particularmente útil em situações de alta demanda como pandemias. O trabalho demonstra como a integração de métodos avançados de IA pode melhorar significativamente a detecção de doenças em imagens médicas, fornecendo uma solução robusta e eficaz para o diagnóstico de COVID-19.

O artigo de KERMANY et al. (2018) explora como técnicas de aprendizado profundo *deep*

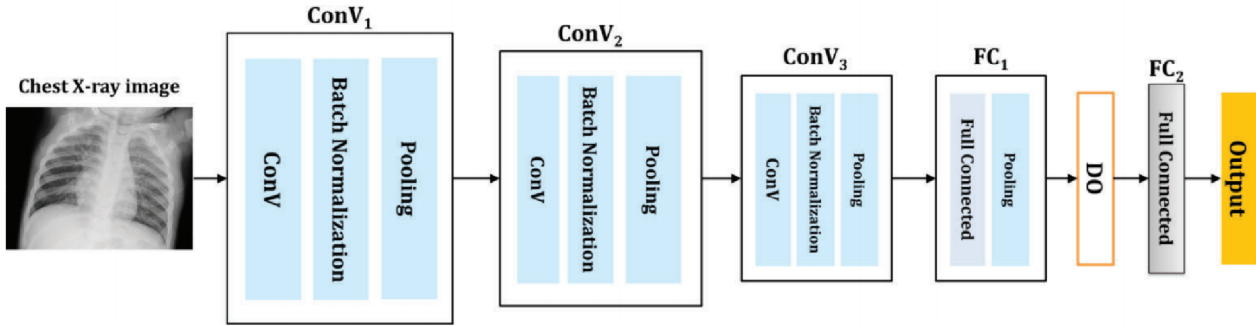


Figura 14: Rede Neural Artificial proposta por REZAEI; BADIEI; MESGHINI (2020).

learning podem ser aplicadas para identificar diagnósticos médicos e doenças tratáveis a partir de imagens médicas. O estudo visa demonstrar como algoritmos de aprendizado profundo, conforme Figura 15, particularmente redes neurais convolucionais (CNNs), podem ser utilizados para analisar imagens médicas, como radiografias, tomografias e ressonâncias magnéticas, para identificar doenças e condições médicas com alta precisão. O treinamento de redes neurais convolucionais em grandes conjuntos de dados de imagens médicas, é realizado com imagens rotuladas, onde cada imagem já é associada a um diagnóstico específico.

O modelo aprende a reconhecer padrões e características associadas a diferentes condições médicas. O estudo demonstra que os modelos de aprendizado profundo podem alcançar ou até superar a precisão dos radiologistas humanos em várias tarefas de diagnóstico, incluindo a identificação de condições como pneumonia, tuberculose e outras doenças pulmonares. O desempenho dos modelos é avaliado por meio de métricas como acurácia, sensibilidade e especificidade.

O uso de aprendizado profundo em imagens médicas pode potencialmente aumentar a eficiência e a precisão no diagnóstico de doenças, além de ajudar na triagem e na detecção precoce de condições tratáveis. Isso pode levar a uma melhoria na gestão e no tratamento dos pacientes. O artigo também discute alguns desafios, como a necessidade de grandes volumes de dados rotulados para treinamento, a variação nas imagens devido a diferentes equipamentos e técnicas de imagem, e a necessidade de validação clínica rigorosa antes da implementação prática. O estudo destaca o potencial do aprendizado profundo para transformar a prática médica, mas também sublinha a importância de uma integração cuidadosa e contínua da tecnologia com a prática clínica.

O estudo de AYAN; ÜNVER (2019) explora a utilização de técnicas de aprendizado profundo (deep learning) para diagnosticar pneumonia com base em imagens de raios-x do tórax. Utilizando um conjunto de dados de imagens de tórax que inclui imagens rotuladas com diferentes condições, como pneumonia e imagens saudáveis (normais). O conjunto de dados é dividido em conjuntos de treinamento, validação e teste para treinar e avaliar o modelo de aprendizado profundo Xception, conforme a Figura 16. Os resultados do estudo indicam que o modelo de aprendizado profundo desenvolvido conseguiu identificar pneumonia a partir das imagens de raios-x com alta precisão. A comparação com métodos tradicionais mostrou que o modelo ba-

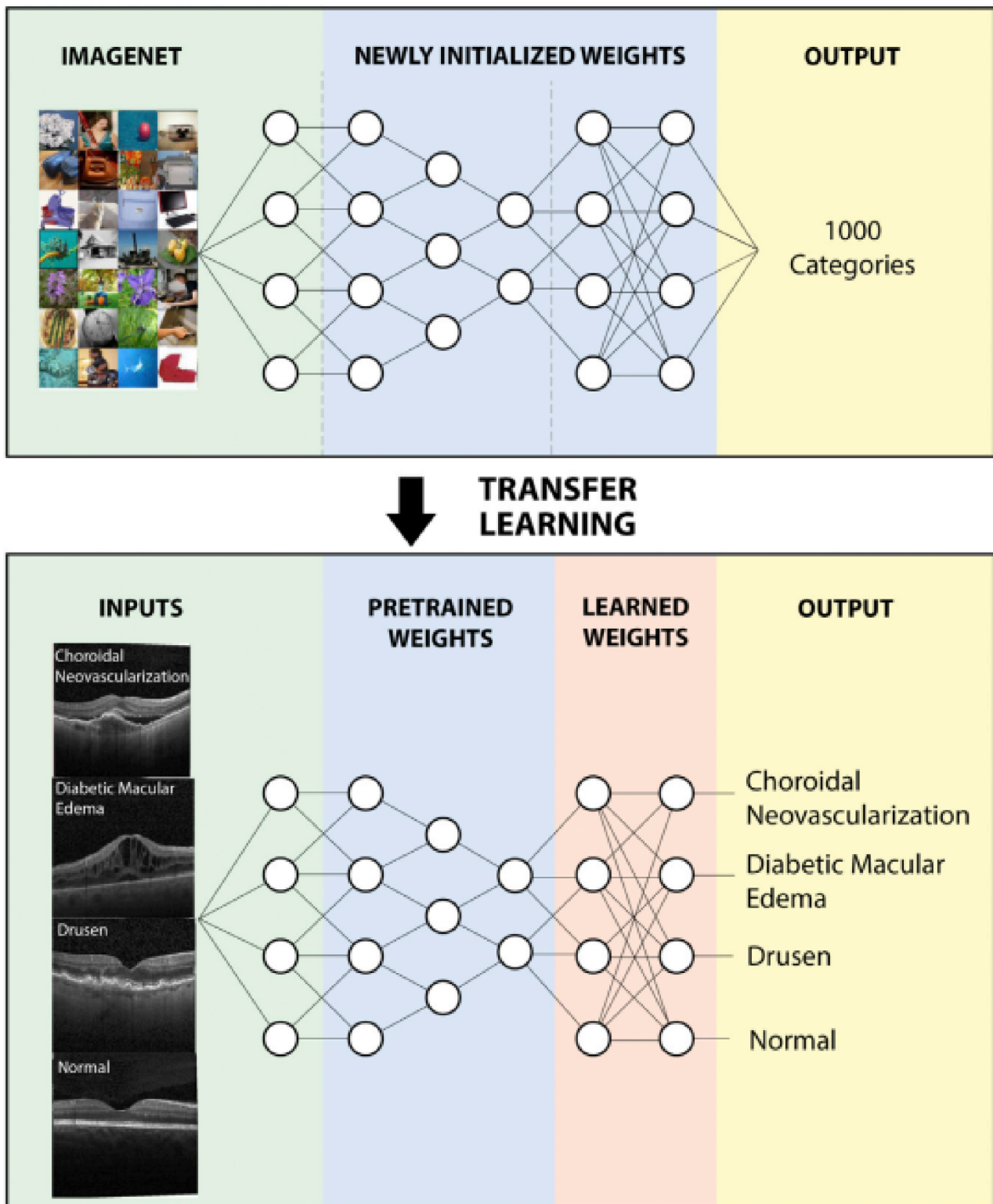


Figura 15: Rede Neural Artificial proposta por KERMANY et al. (2018).

seado em deep learning pode oferecer uma precisão semelhante ou superior, reduzindo a taxa de erros e o tempo necessário para o diagnóstico.

Para os profissionais de diagnóstico por imagens, é necessário o desenvolvimento das novas tecnologias de auxílio ao diagnóstico. A incorporação da visão computacional como uma fer-

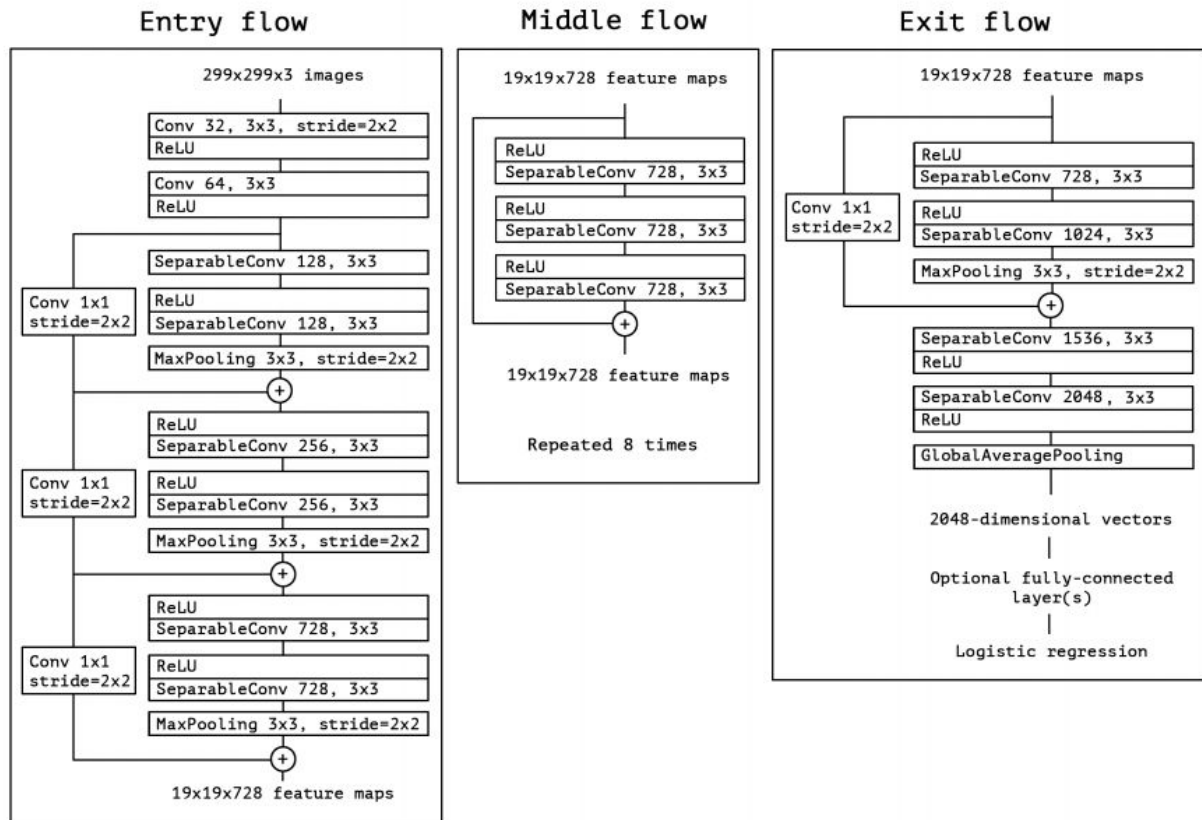


Figura 16: Rede Neural Artificial utilizada por AYAN; ÜNVER (2019).

ramenta é essencial por sua adaptabilidade e interdisciplinariedade com todos os processos de filtragem e processamento das imagens médicas.

3 METODOLOGIA

Este capítulo tem como intuito descrever a metodologia que será utilizada, considerando o desenvolvimento da modelagem da rede neural para o treinamento e classificação dos padrões.

3.1 *Dataset*

A amostra de imagens selecionadas foi extraída de domínio público KERMANY; ZHANG; GOLDBAUM (2018), sendo referenciada no artigo *Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning* (KERMANY et al., 2018).

O uso de amostras de imagens médicas de domínio público desempenha um papel crucial no avanço da pesquisa em saúde e medicina. Essas amostras são geralmente disponibilizadas de forma aberta e gratuita, promovendo a colaboração, inovação e desenvolvimento de algoritmos de aprendizado de máquina aplicados à análise de imagens médicas. Alguns pontos importantes sobre o uso de amostras de imagens médicas de domínio público são o acesso livre e a colaboração das imagens médicas de domínio público, que estão disponíveis para pesquisadores, acadêmicos e desenvolvedores sem restrições, facilitando o acesso e promovendo a colaboração global. Isso acelera a pesquisa e o desenvolvimento de soluções inovadoras (BRANCO, 2011).

O treinamento de modelos: Essas amostras são frequentemente usadas para treinar modelos de aprendizado de máquina, especialmente em tarefas como detecção de doenças, segmentação de órgãos e classificação de condições médicas. Ter conjuntos de dados públicos ajuda a melhorar a generalização e a eficácia dos modelos. Validação e *Benchmarking* são o conjunto de dados de imagens médicas de domínio público que servem como *Benchmarking* padrão para avaliar o desempenho de algoritmos. Isso permite a comparação justa entre diferentes abordagens e impulsiona a qualidade das soluções propostas.

Desenvolvimento de Novos Métodos: Pesquisadores utilizam esses conjuntos de dados para desenvolver e aprimorar métodos de processamento de imagem e aprendizado de máquina específicos para a área médica. Isso inclui a automação de diagnósticos, análises de padrões e identificações de biomarcadores (BRANCO, 2011).

O uso de amostras de domínio público aumenta a transparência e a reprodutibilidade da pesquisa, uma vez que outros pesquisadores podem acessar os mesmos dados para validar resultados e construir sobre trabalhos anteriores. A ética e a privacidade ao utilizar amostras de imagens médicas de domínio público é essencial para garantir que questões sejam consideradas. A anonimização apropriada dos dados é crucial para proteger a identidade dos pacientes (BRANCO, 2011).

3.2 Seleção de amostras

Os dados para as análises do presente trabalho foram utilizados imagens de raios-x da região torácica, essas imagens são utilizadas para diversos diagnósticos. O complexo torácico envolve

uma gama de tecidos de diferentes composições, os tecidos de interesse para esse trabalho são os tecidos pulmonares. Com essa secção de corte, trabalhamos com 3 classes de imagens: as normais, com pneumonia bacteriana e pneumonia viral.

Na Figura 17, podemos observar que as classes de imagens são separadas em pastas no diretório, assim constituindo o *Dataset*.

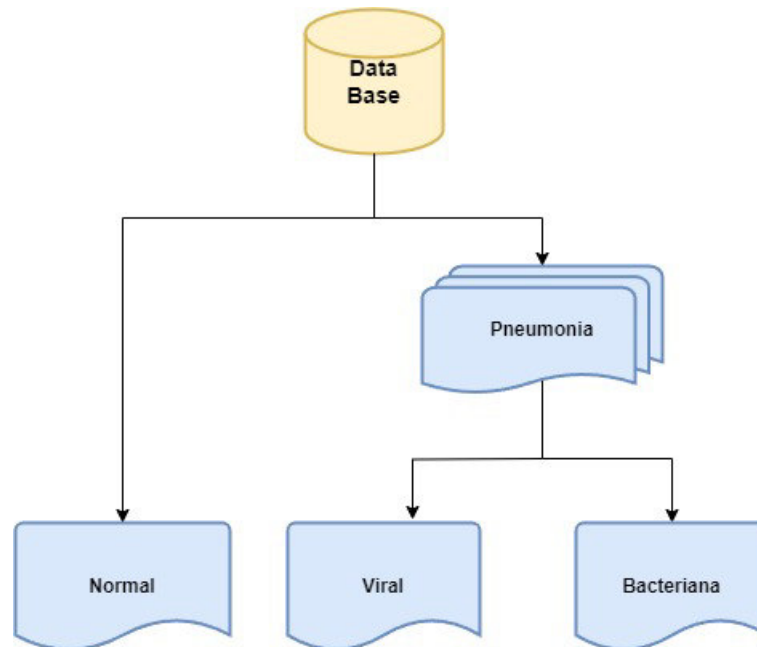


Figura 17: *Dataset*.

A separação do *Dataset* em três classes evidencia a importância da identificação dos tipos de pneumonia, essa separação, de duas classes para três aumenta a complexidade do treinamento da rede neural de classificar os padrões das imagens.

3.3 Treinamento da rede neural

A validação do modelo VGG para treinamento de imagens médicas foi realizada utilizando os pacotes *Keras* e *TensorFlow*. Foram escolhidos porque são uma biblioteca de alto nível para *deep learning* e com várias bibliotecas para treinamento de redes neurais. A linguagem principal do código é escrita em *Python*.

O treinamento da rede segue a seguinte sequência de evolução: validação do modelo base, treinamento com adição de camadas ocultas, teste e validação.

Usando o modelo VGG como base para a identificação dos padrões já aprendidos no *imagenet/VGG*, adicionaremos os nossos novos dados para imagens pulmonares de nosso banco de dados para o treinamento e aprendizagem dos novos padrões.

Quando chamamos o modelo VGG, adicionamos novas camadas densas para a identificação das imagens pulmonares e solicitamos a geração de um treinamento com o novo banco de dados.

São introduzidas três classes para entrada do sistema, o tamanho das imagens 224×224 e o tamanho do lote de 32 imagens.

Após a separação dos dados, geramos o *plot* de geração aleatória de treinamento com as imagens do banco de dados conforme a Figura 18.

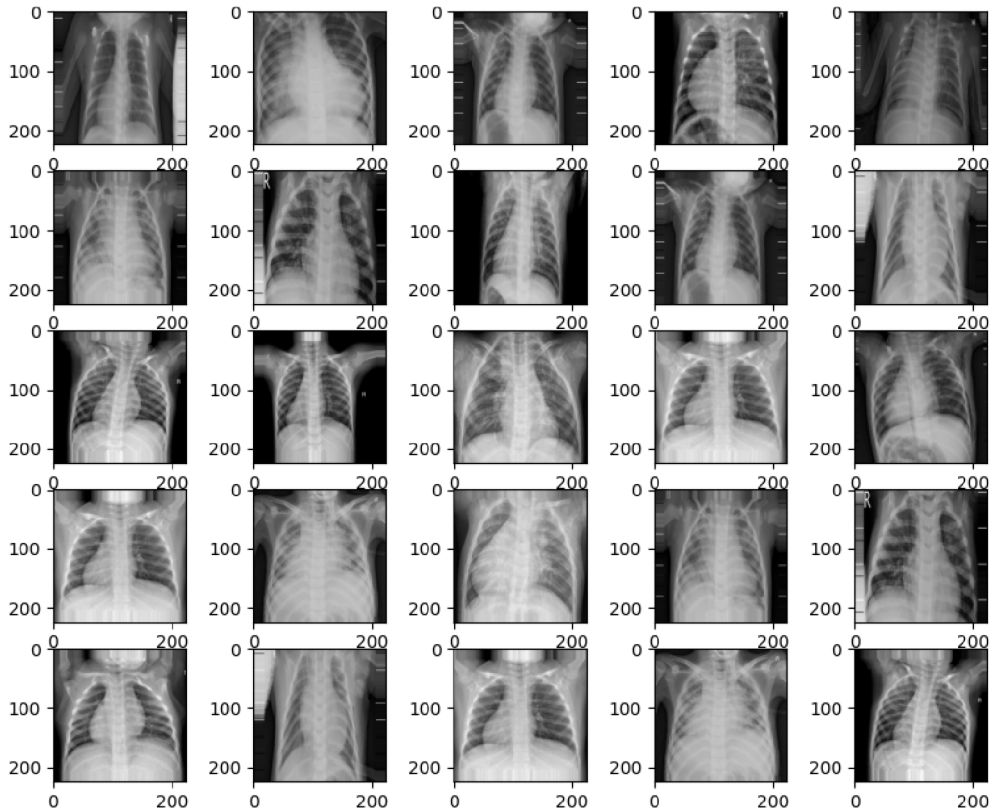


Figura 18: Conjunto parcial de dados gerado pelo treinamento de rede neural.

Fazendo o carregamento do modelo VGG e compilando o modelo gerado para aprendizagem em 10 *epoch* em tradução direta épocas. Gera-se o gráfico de perda de aprendizagem por época conforme a Figura 19.

Após o treinamento, é gerado um teste de verificação da comparação dos padrões aprendidos com as imagens dos bancos de dados para teste e validação do modelo, conforme a Figura 20. Essa plotagem é gerada com os dados sem a diferenciação dos tipos de pneumonia, o modelo é válido para a diferenciação dos padrões de imagem com e sem pneumonias.

O treinamento detalhado com base no modelo VGG foi detalhado no fluxograma da Figura 21, o modelo tem como entrada *Input* o banco de dados com as classes separadas como: Normal, Bactéria e Vírus, em sequência passa para um modelo pré-treinado o VGG, em que os parâmetros são congelados e adicionamos outras camadas densas, compostas por uma camada densa de 1024 neurônios, um *batch* de normalização, mais uma camada densa com 1024 neurônios, uma camada densa com 3 neurônios e uma saída *output*.

A estrutura da rede é composta por duas partes principais, a base pré-treinada e a ca-

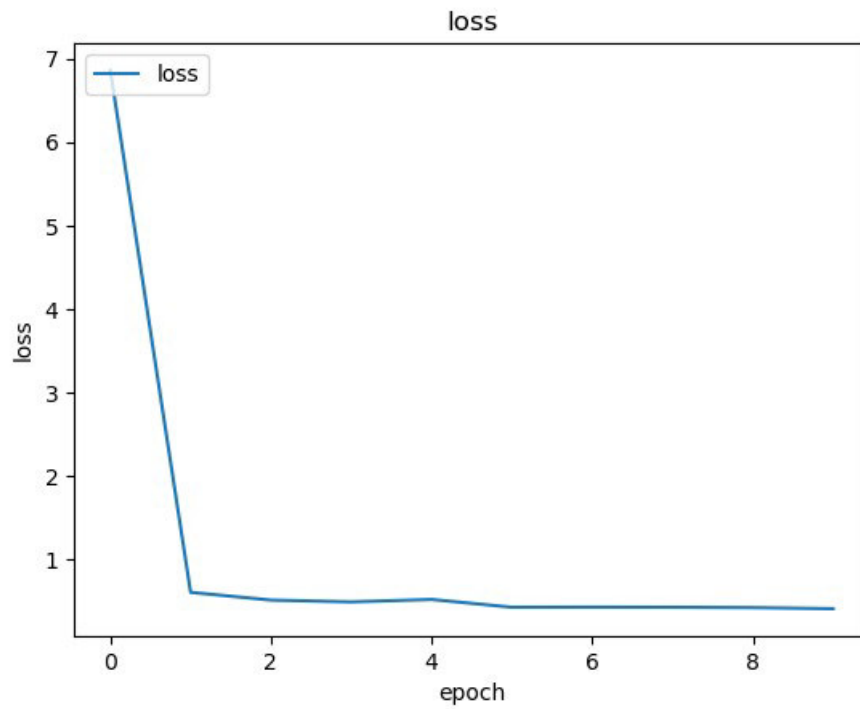


Figura 19: Gráfico de perda, no treinamento da rede em 10 épocas.

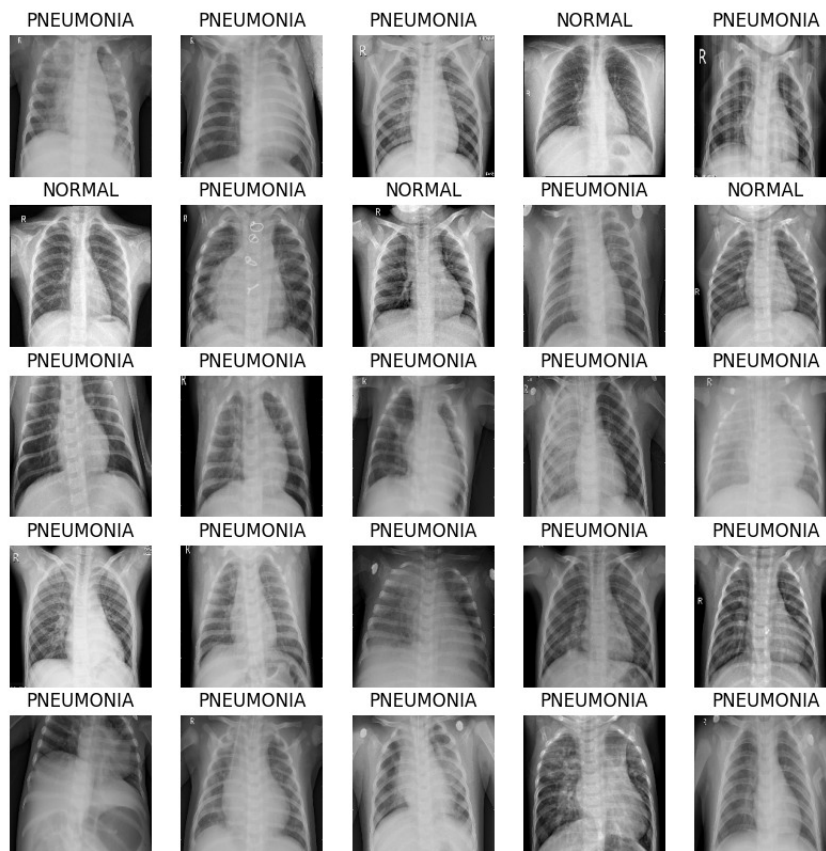


Figura 20: Conjunto de dados do treinamento do modelo VGG.
Fonte: Autor

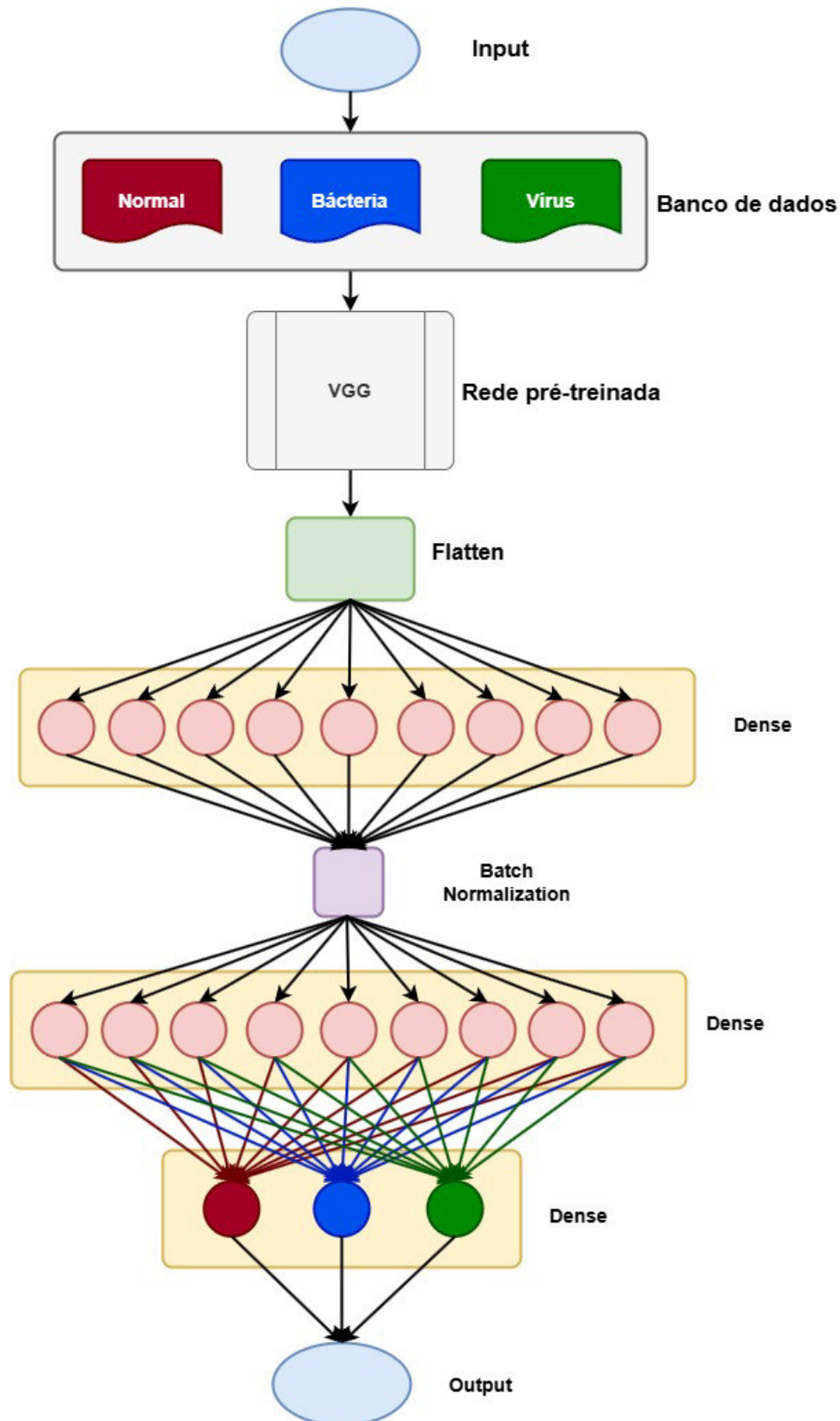


Figura 21: Fluxograma do modelo treinamento da rede neural

mada personalizada. A base pré-treinada utiliza a arquitetura VGG16 como extrator de características. As camadas convolucionais e de *pooling* desta rede foram treinadas no *dataset ImageNet* e já conhecem padrões visuais para diversas classificações e diferenciações. A ca-

mada personalizada são camadas densas adicionadas no topo da VGG16 para realizar a tarefa de classificação específica (imagens de pneumonia).

3.4 Componentes da Rede Neural

Neste subcapítulo, descrevemos a rede neural e seus componentes.

O modelo base para o pré-treinamento, o modelo VGG, conforme discutido no capítulo de Estado da Arte, possui parâmetros para treinamento para vários tipos de classificação de imagens. A rede é congelada no final da rede VGG para manter os parâmetros e os pesos da rede. Então, adicionamos as novas camadas ocultas.

A base VGG16 é inicializada com pesos pré-treinados no *dataset ImageNet*.

```
1 base_model = VGG16(weights='imagenet', include_top=False,
2     input_shape=(image_size, image_size, 3))
```

As camadas densas originais da VGG16 são removidas, isso permite a adição das nossas próprias camadas densas para uma tarefa específica, a classificação das imagens de pneumonia Bacteriana e Viral.

```
1 include_top=False:
```

O *input_shape* define o tamanho das imagens de entrada (224 x 224 pixels com 3 canais RGB).

```
1 input_shape=(image_size, image_size, 3):
```

No final da rede VGG16, congelamos as camadas ocultas da rede.

Congelamento de Camadas

```
1 for layer in base_model.layers:
2     layer.trainable = False
```

Todas as camadas da VGG16 foram congeladas, ou seja, para que seus pesos não atualizem durante o treinamento. Isso evita o desaprendizado dos padrões já conhecidos.

A camada personalizada foi composta pela adição das novas camadas responsáveis pelas diferenciações dos novos padrões de classes: normal, bacteriana e viral. Ou seja, a camada personalizada é composta pelas camadas densas adicionadas: *Flatten*, *batch normalization*, duas de *dense* e uma de saída.

O *Flatten* foi responsável pela conversão do tensor 3D da saída da VGG16 em um vetor 1D para ser processado pelas camadas densas.

O *Batch normalization* normaliza a saída para acelerar o treinamento e estabilizar as ativações.

Já as camadas *dense* foram compostas por 1024 neurônios, sendo responsáveis pelo aprendizado dos novos padrões.

A última camada *dense* possui 3 neurônios, pois a probabilidade gerada para saída foi igual à quantidade de classes de entrada.

Camada Personalizada

```

1
2 model = Sequential()
3     model.add(base_model)
4     model.add(Flatten())
5     model.add(Dense(1024, activation='relu'))
6     model.add(BatchNormalization())
7     model.add(Dense(1024, activation='relu'))
8     model.add(Dense(num_class, activation='softmax'))
9     optimizer = SGD(learning_rate=0.001)

```

A função de perda e métricas são responsáveis por gerar as avaliações de aprendizado do treinamento e validação da rede neural. A função de ativação ReLU (*Rectified Linear Unit*) é amplamente utilizada em redes neurais devido à sua simplicidade e eficácia. Ela transforma a entrada x de acordo com a fórmula $f(x) = \max(0, x)$, ou seja, valores negativos são mapeados para zero, enquanto valores positivos permanecem inalterados. Isso permite que o modelo aprenda representações complexas ao introduzir não-linearidades na rede, essencial para resolver problemas não-lineares. Além disso, a ReLU é computacionalmente eficiente, ao evitar operações complexas, como exponenciais ou sigmóides. No entanto, apresenta algumas limitações, como o problema de neurônios mortos, onde alguns neurônios podem parar de aprender se os gradientes se tornarem constantemente zero.

Função de Perda e Métricas

```

1 model.compile(
2     optimizer=optimizer,
3     loss='categorical_crossentropy',
4     metrics=['categorical_accuracy', 'Precision', 'Recall', 'binary_accuracy']
5 )

```

A augmentação tem o papel crucial no treinamento, pois sua função é aumentar a quantidade de dados, como: rotacionar a imagem, fazer a magnificação, a modificação do alcance do cisalhamento e a distorção da imagem. Dessa forma, ele gera dados para o treinamento com divisão para validação, logo, ele aumenta a quantidade de dados para o treinamento e separa o percentual de 30% dos dados para validação do modelo.

Augmentação de Dados no Treinamento

```

1 train_datagen = ImageDataGenerator(
2     validation_split=0.3,
3     shear_range=0.2,
4     zoom_range=0.2,
5     horizontal_flip=True,
6     rescale=1./255
7 )

```

O gerador de treinamento é a função no código que utiliza da augmentação junto ao banco de dados, configurando a entrada do tamanho das imagens para *image size* 224×224 e o tamanho do *batch size* em tradução literal (tamanho do pacote), sendo ele composto por 32

imagens, processadas em 108 etapas de cada época do treinamento.

Geradores de Treinamento e Validação

```
1 train_generator = train_datagen.flow_from_directory(
2     directory=train_path,
3     target_size=(image_size, image_size),
4     batch_size=32,
5     class_mode='categorical',
6     subset='training')
```

A execução do treinamento possui um processo de parada antecipada caso o *early stopping* constate uma repetição de 5 valores iguais para a variável *loss* (perda de aprendizado) nos parâmetros do treinamento, assim *model checkpoint* salva o modelo e encerra o treinamento.

Treinamento

```
1 early_stopping = EarlyStopping(monitor='val_loss', patience=5,
2     restore_best_weights=True)
3
4 model_checkpoint = ModelCheckpoint('best_model.keras', save_best_only=True)
5
6 history = classification_model.fit(
7     train_generator,
8     epochs=50, # 50 epochs
9     validation_data=validation_generator,
10    callbacks=[early_stopping, model_checkpoint])
```

O modelo é treinado por 50 épocas. Durante cada época, ele avalia a perda e as métricas no conjunto de validação.

Após o treinamento e validação, é gerado um teste com novas imagens inéditas ao sistema para verificar a acurácia dos padrões aprendidos.

Teste no Conjunto de Teste

```
1 loss, accuracy, precision, recall, binary_accuracy = classification_model.
2     evaluate(test_generator)
```

O desempenho final é avaliado comparando os parâmetros de treinamento, validação e teste.

3.5 Métricas de avaliação do modelo

A matriz de confusão representada na Tabela 1 é uma ferramenta essencial para avaliar o desempenho de modelos de classificação, especialmente em problemas de classificação binária e multiclasse. Ela fornece uma visualização clara de como as previsões do modelo se comparam aos valores reais.

Em um problema de classificação com C classes, a matriz de confusão $\mathbf{M}$ é uma matriz quadrada de dimensão $C \times C$, onde o elemento M_{ij} representa o número de amostras da classe real i que foram classificadas como pertencentes à classe j . Dessa forma, a diagonal principal M_{ii} contém as previsões corretas (*True Positives* - TP), enquanto os elementos fora da diagonal representam erros de classificação.

Tabela 1: Matriz de Confusão.

	Prevista 1	Prevista 2	Prevista 3
Real 1	(TP)	(FN)	(FN)
Real 2	(FP)	(TP)	(FN)
Real 3	(FP)	(FP)	(TP)

Para um problema de classificação binária, a matriz de confusão é geralmente apresentada em uma forma 2×2 , para o nosso modelo de multiclass a matriz de confusão é apresentada 3×3 . A matriz 3×3 apresentada na Tabela 1, é descrita na lateral direita pelo termo **Real 1, 2 e 3** que corresponde as verdadeiras classes separadas no *Dataset* usadas para treinar a rede neural e as preditas que correspondem pelas predições do modelo apos o aprendizado dos padrões e classes. As definições dos termos que compões a matriz de confusão são descritos como:

- *True Positive (TP)* Verdadeiro Positivo: O modelo previu positivo e o valor real também é positivo.
- *False Positive (FP)* Falso Positivo: O modelo previu positivo, mas o valor real é negativo.
- *False Negative (FN)* Falso Negativo: O modelo previu negativo, mas o valor real é positivo.
- *True Negative (TN)* Verdadeiro Negativo: O modelo previu negativo e o valor real também é negativo.

Em uma matriz de confusão para problemas de múltiplas classes, a análise pode ser feita como uma matriz binaria de duas classes como representada na Figura 22, as linhas mostram as classes preditas e as colunas representam as classes reais. A diagonal da matriz contém os valores corretos (TP), enquanto os valores fora da diagonal representam erros de classificação (FP e FN).

A Matriz de Confusão é importante porque fornece uma visualização das predições de modelo, essa visualização é uma representação clara de onde o modelo está se saindo bem e onde está cometendo erros. A identificação de erros ajuda a entender se o modelo tende a classificar certos tipos de exemplos incorretamente e as informações da matriz podem orientar ajustes no modelo, como a escolha de limiares ou técnicas de balanceamento de classes.

3.5.1 Métricas Derivadas da Matriz de Confusão

A matriz de confusão permite calcular várias métricas de desempenho as equações a seguir apresentam as equações correspondentes as métricas de acurácia, presisão, *recall* e *F1-score*.

$$\text{Acurácia} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.1)$$

		True Class		
		Class 1	Class 2	Class 3
Predicted Class	Class 1	TP	FP	FP
	Class 2	FN	TN	TN
	Class 3	FN	TN	TN

Figura 22: Matriz genérica para classificação multiclasse

$$\text{Precisão} = \frac{TP}{TP + FP} \quad (3.2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.3)$$

$$F1 = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (3.4)$$

As métricas de acurácia, precisão, *recall* e *F1-score* são fundamentais na avaliação de modelos de machine learning, especialmente em problemas de classificação. A acurácia mede a proporção de previsões corretas em relação ao total, sendo útil quando as classes estão balanceadas. A precisão indica a taxa de exemplos corretamente classificados como positivos entre todos os previstos como positivos, essencial quando os falsos positivos devem ser minimizados. O *recall* mede a capacidade do modelo de identificar corretamente os exemplos positivos, sendo crucial em cenários onde falsos negativos são mais problemáticos. O *F1-score* é a média harmônica entre precisão e *recall*, proporcionando um equilíbrio entre esses dois aspectos, especialmente útil quando há um desequilíbrio entre as classes. Essas métricas, em conjunto, oferecem uma visão abrangente do desempenho do modelo, permitindo ajustes e melhorias na tomada de decisão.

Outras métricas usadas para a avaliação do desempenho geral do modelo são o *loss*, *Categorical Crossentropy*, *Binary Crossentropy*.

A definição de *loss* ou função de perda é a medida da diferença entre as previsões do modelo e os valores reais. É uma medida da qualidade do modelo, quanto menor a perda, melhor o

modelo. Essa métrica é usada durante o treinamento, o objetivo é minimizar a função de perda ajustando os pesos do modelo. Várias funções de perda podem ser usadas, dependendo do tipo de problema, podendo ser uma Regressão (*Mean Squared Error*) (MSE) ou Classificação por *Categorical Crossentropy* para múltiplas classes ou *Binary Crossentropy* para duas classes.

A definição de *categorical accuracy* é a medição da proporção de previsões corretas do modelo em relação ao total de amostras, conforme mostrada na Equação 3.5. É uma métrica apropriada para problemas de classificação com múltiplas classes.

$$\text{categorical acurácia} = \frac{\text{Número de previsões corretas}}{\text{Número total de amostras}} \quad (3.5)$$

A *binary accuracy* é uma métrica de avaliação usada em problemas de classificação binária. Ela calcula a precisão do modelo, ou seja, a proporção de previsões corretas feitas em relação ao total de previsões. Simplificando, a métrica mede quantas vezes o modelo classificou corretamente as amostras como sendo de uma das duas classes possíveis (geralmente 0 ou 1).

Em classificação binária, o modelo faz previsões para duas classes (0 ou 1, por exemplo). A precisão binária é calculada comparando as previsões do modelo com os valores reais das classes, conforme as Equações 3.6 e 3.7.

$$\text{binary accuracy} = \frac{\text{Número de previsões corretas}}{\text{Número total de previsões}} \quad (3.6)$$

Ou seja:

$$\text{binary accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.7)$$

Dessa forma:

- TP (True Positive): Previsão correta de um exemplo da classe positiva (1).
- TN (True Negative): Previsão correta de um exemplo da classe negativa (0).
- FP (False Positive): Previsão incorreta de um exemplo da classe negativa como sendo da classe positiva.
- FN (False Negative): Previsão incorreta de um exemplo da classe positiva como sendo da classe negativa.

Como é calculada no *Keras* e *TensorFlow*: Em *frameworks* como *Keras*, *binary accuracy* funciona assim, o modelo gera uma probabilidade de uma amostra pertencer à classe positiva (1). Para a métrica de acurácia binária, as previsões de probabilidade são arredondadas para 0 ou 1, com base em um limiar (geralmente 0.5).

- Se a probabilidade prevista for maior ou igual a 0.5 o modelo prevê 1 (classe positiva).
- Se for menor que 0.5 o modelo prevê 0 (classe negativa).

Depois disso, é feita uma comparação entre essas previsões binárias (0 ou 1) e os rótulos verdadeiros para calcular a proporção de acertos.

A *binary accuracy* é normalmente usada em tarefas de classificação binária, como:

- Classificação de *e-mails* (spam ou não spam).
- Diagnóstico de doenças (presença ou ausência de uma doença).
- Detecção de fraudes (fraude ou não fraude).

A diferença entre *binary accuracy* e outras métricas como *categorical accuracy* que é usada para problemas de classificação com mais de duas classes (classificação multcategórica) ou a *sparse categorical accuracy* em tradução direta (precisão categórica esparsa) que mede a precisão em problemas de classificação multiclasse, comparando a classe prevista com a classe verdadeira representada como um inteiro em vez de um vetor *one-hot*, mas usada quando os rótulos verdadeiros estão em formato de inteiros (em vez de um vetor de probabilidades).

A *binary accuracy* é uma métrica essencial quando o problema tem apenas duas classes, e ela mede quão bem o modelo faz previsões corretas nessas duas classes.

Curva ROC e AUC

A curva ROC do inglês (Receiver Operating Characteristic) (ROC) em tradução direta (Característica operacional do receptor) é uma ferramenta gráfica usada para avaliar o desempenho de um modelo de classificação binária. Para traçar a curva ROC a partir da matriz de confusão, é necessário calcular as taxas de verdadeiros positivos (TPR) e falsos positivos (FPR) para diferentes limiares de decisão.

Como temos uma matriz de confusão que representa a classificação de múltiplas classes (Normal, Bactéria e Vírus), precisamos seguir um processo que envolve tratar cada classe como um problema binário (um contra todos).

A curva ROC demonstra a relação entre dois indicadores principais:

A (TPR) também conhecida como **Sensibilidade**, que é a proporção de positivos corretamente classificados pelo modelo em relação ao total de positivos reais demonstrado na Equação 3.8.

$$TPR = TP / (TP + FN) \quad (3.8)$$

A (FPR) também conhecida como **Especificidade**, que corresponde a proporção de negativos incorretamente classificados como positivos em relação ao total de negativos reais mostrado na Equação 3.9.

$$FPR = FP / (FP + TN) \quad (3.9)$$

Em que FP são falsos positivos e TN são verdadeiros negativos.

A construção da Curva ROC é feita plotando a Taxa de Verdadeiros Positivos (TPR) contra a Taxa de Falsos Positivos (FPR) para diferentes valores de limiar de decisão. O limiar de decisão é o valor que define a partir de qual probabilidade o modelo decide entre as classes.

Em cada ponto da curva, o modelo classifica as instâncias com base em um determinado limiar. Quando o limiar é ajustado para diferentes valores, o TPR e o FPR também mudam, criando uma curva que reflete o comportamento do modelo.

Interpretação da Curva ROC:

Canto superior esquerdo da curva (ponto 0,1) representa a classificação ideal, em que a Taxa de Falsos Positivos (FPR) é zero e a Taxa de Verdadeiros Positivos (TPR) é 1. Diagonal de 45 graus (linha pontilhada entre os pontos 0,0 e 1,1) representa um classificador aleatório, ou seja, um modelo sem poder preditivo. Quanto mais a curva se aproxima do canto superior esquerdo e se distancia da diagonal de 45 graus, melhor é o modelo.

(*Area Under the Curve*) (AUC) (Área Sob a Curva):

A AUC do inglês (*Area Under the Curve*) (AUC) em tradução direta (Área Sob a Curva) é um número entre 0 e 1 que mede a Área Sob a Curva ROC, é uma métrica que quantifica a capacidade de um modelo em distinguir entre as classes positivas e negativas. Ela é calculada a partir da curva ROC, que plota a taxa de verdadeiros positivos (TPR) contra a taxa de falsos positivos (FPR) em diferentes limiares de decisão.

A AUC pode ser calculada utilizando:

$$AUC = \sum_{i=1}^{n-1} \frac{1}{2} \times (FPR_{i+1} - FPR_i) \times (TPR_{i+1} + TPR_i) \quad (3.10)$$

Onde:

- FPR_i e TPR_i São valores para TPR e FPR no ponto i .
- FPR_{i+1} e TPR_{i+1} São os valores de FPR e TPR para o próximo ponto.

O princípio da regra do trapézio é entre dois pontos consecutivos, a área é aproximada pela média das alturas dos pontos TPR_i e TPR_{i+1} multiplicada pela largura (a diferença entre os valores de FPR).

A AUC varia entre 0 e 1, sendo que um valor de 1 indica um modelo perfeito que consegue separar totalmente as classes, enquanto um valor de 0,5 indica que o modelo tem o desempenho de uma classificação aleatória. Quanto maior a AUC, melhor é o desempenho do modelo em termos de discriminação entre as classes. Ela fornece uma visão geral do desempenho do modelo, independentemente do limiar de classificação escolhido, sendo útil em problemas de classificação desbalanceada.

- $AUC = 1$ Indica um classificador perfeito.
- $AUC = 0,5$ Modelo aleatório, sem poder preditivo.
- $AUC \geq 0,9$ Excelente capacidade preditiva.
- $0,7 \leq AUC \leq 0,9$ Bom desempenho, mas pode haver espaço para melhorias.

4 RESULTADOS E DISCUSSÕES

Neste capítulo, estaremos apresentando e analisando e discutindo os resultados.

4.1 *Dataset*

A separação do *Dataset* resultou em 3 classes conforme a Figura 23. A diferenciação do banco de dados é importante para o treinamento e aprendizagem dos padrões de cada classe de imagem.

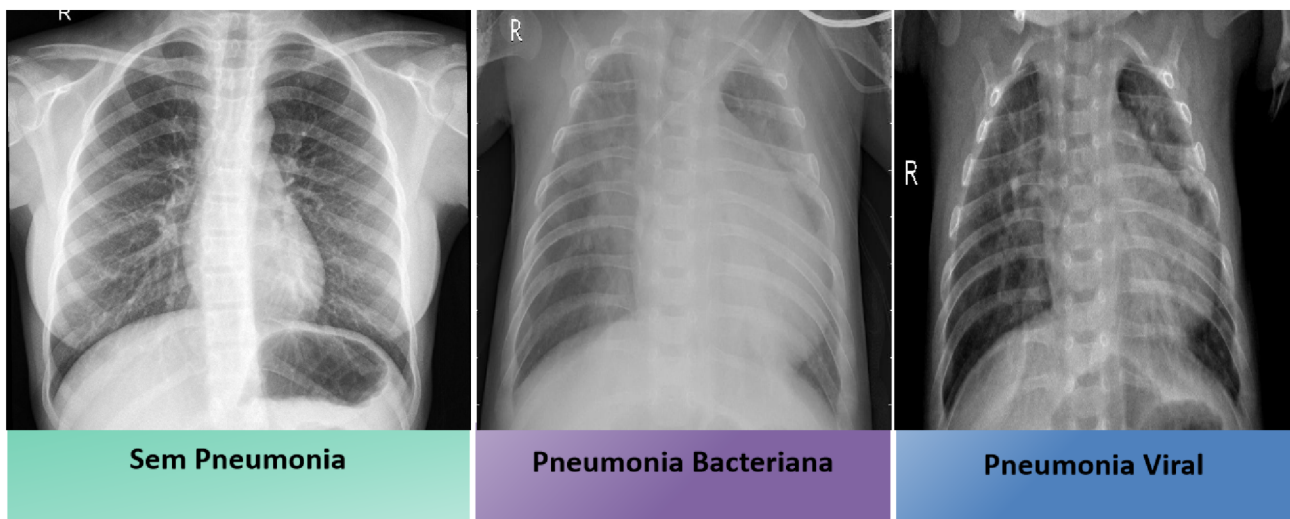


Figura 23: Exemplo da separação das classes.

A abordagem de manter a separação dos dados até o final dos resultados tem como premissa a importância de ressaltar a diferenciação dos tipos de pneumonia, como forma de auxílio ao diagnóstico médico.

Foram utilizadas 5.527 imagens de raio-x de tórax na rede neural, sendo 1.278 para sem pneumonia (normal), 1.810 para pneumonia bacteriana e 1.816 para pneumonia viral.

Tabela 2: Distribuição do Banco de Dados para o Modelo

	Treinamento	Validação	Teste
Normal	1.278	383	235
Bactéria	1.810	543	242
Vírus	1.816	544	148
Total	4.904	1.470	625

O *Dataset* usado estava disponível e com domínio publico para uso pelo KERMANY; ZHANG; GOLDBAUM (2018), os dados estavam apenas separados como normal e pneumonia, para os dados sem pneumonia eram chamados de normal e as pneumonias bacterianas e virais

estavam em uma única pasta. A disparidade nas quantidades das imagens para teste resultou da separação por classes das imagens do *Dataset*.

4.2 Treinamento e Validação do Modelo

Conforme o capítulo de materiais e métodos, o modelo base para a rede neural que utilizamos foi o VGG, em que suas camadas finais foram congeladas para preservar os parâmetros e pesos dessa rede, utilizando-a como uma rede pré-treinada, juntamente com a adição de novas camadas para a classificação das imagens de pneumonia bacteriana e viral.

A Tabela 3 apresenta as métricas obtidas no treinamento e a 4 na validação do modelo.

Tabela 3: Métricas de Treinamento

Epoch	Loss	Categorical accuracy	Precision	Recall	Binary accuracy
1	0.6481	0.7229	0.7638	0.6572	0.818
2	0.5176	0.7875	0.8059	0.7560	0.858
3	0.4718	0.8094	0.8255	0.7886	0.874
4	0.4521	0.8116	0.8280	0.7935	0.8762
5	0.4266	0.8234	0.8350	0.8050	0.882
6	0.4144	0.8308	0.8424	0.8154	0.8876
7	0.4069	0.8335	0.8480	0.8171	0.8902
8	0.4028	0.8368	0.8471	0.8209	0.8909
9	0.3895	0.8401	0.8512	0.8253	0.8937
10	0.3856	0.8401	0.8498	0.8289	0.8941

Tabela 4: Métricas de Validação

Epoch	Val Loss	Val Categorical Accuracy	Val Precision	Val Recall	Val Binary Accuracy
1	0.7772	0.6522	0.7151	0.4942	0.7658
2	0.6629	0.7174	0.7693	0.6439	0.8169
3	0.6149	0.7155	0.7378	0.6835	0.8135
4	0.6361	0.7231	0.7297	0.6905	0.8116
5	0.6756	0.7008	0.7152	0.6777	0.8026
6	0.6889	0.7091	0.7212	0.6963	0.8090
7	0.6642	0.7084	0.7216	0.6861	0.8071
8	0.7139	0.6905	0.6975	0.6650	0.7922
9	0.7351	0.6912	0.6992	0.6733	0.7945
10	0.6448	0.7276	0.7409	0.7168	0.8220

Tabela 5: Métricas de Teste

Epoch	Teste Loss	Teste Categorical Accuracy	Teste Precision	Teste Recall	Teste Binary Accuracy
1	0.3965	0.8816	0.8918	0.8704	0.9216

A partir das métricas obtidas pelo treinamento e validação da rede neural, obtemos os seguintes gráficos representados nas figuras 24 e 25, sendo eles os gráficos *Loss*, *Precision*, *Categorical accuracy* e *Binary accuracy*.

As Figuras 24 e 25 apresentam os gráficos para todas as métricas de treinamento e de validação do modelo, realizadas com 50 épocas. Devido ao *early stopping* o modelo encerrou após 10 épocas.

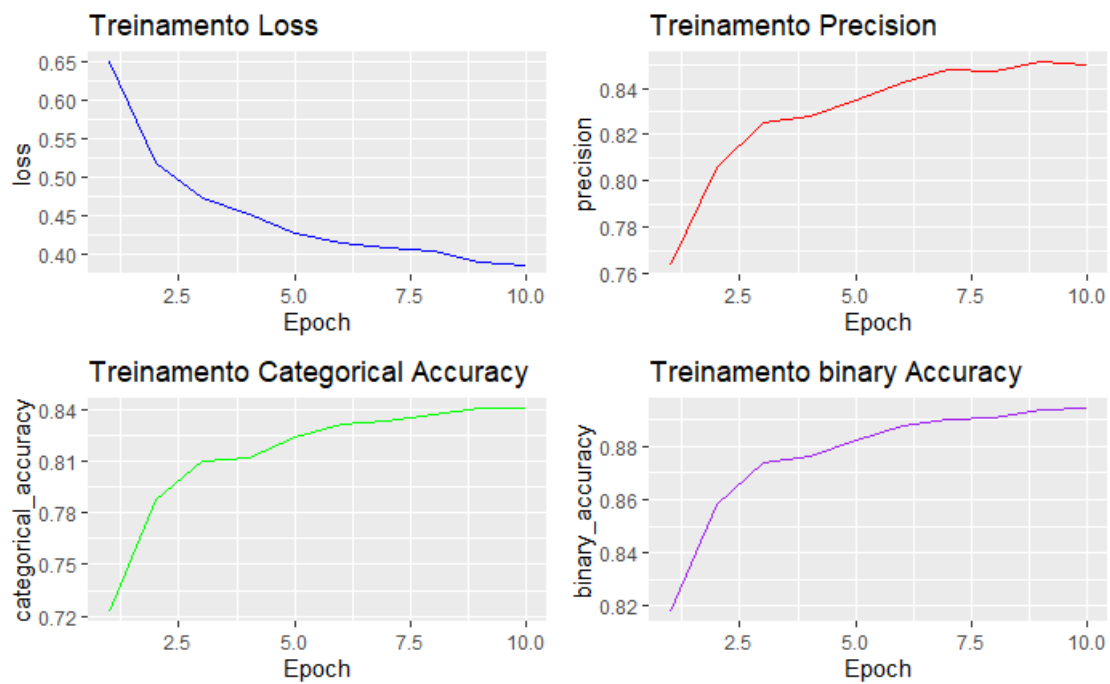


Figura 24: Gráfico das métricas de treinamento do modelo.

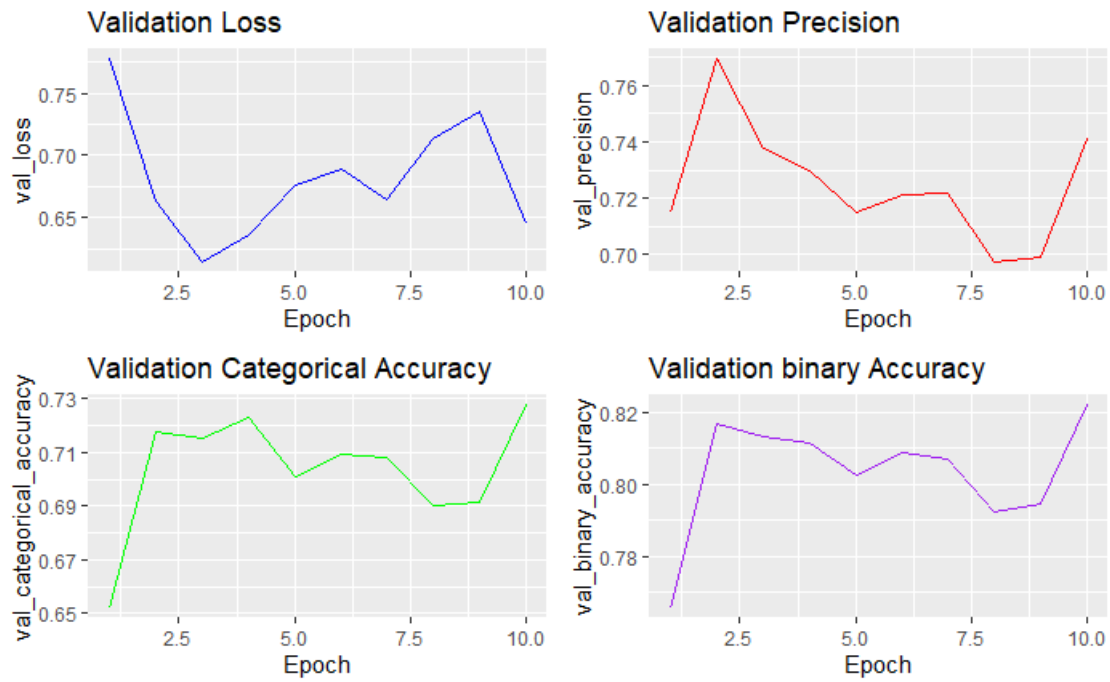
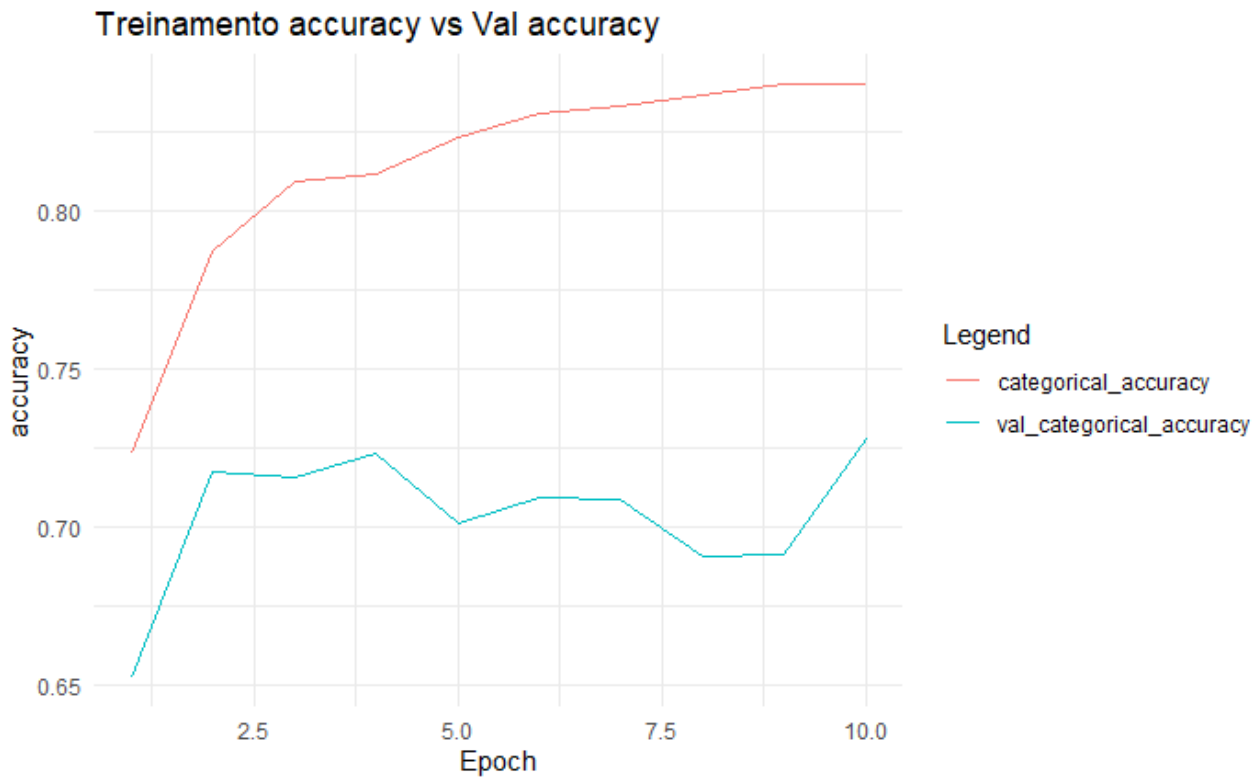
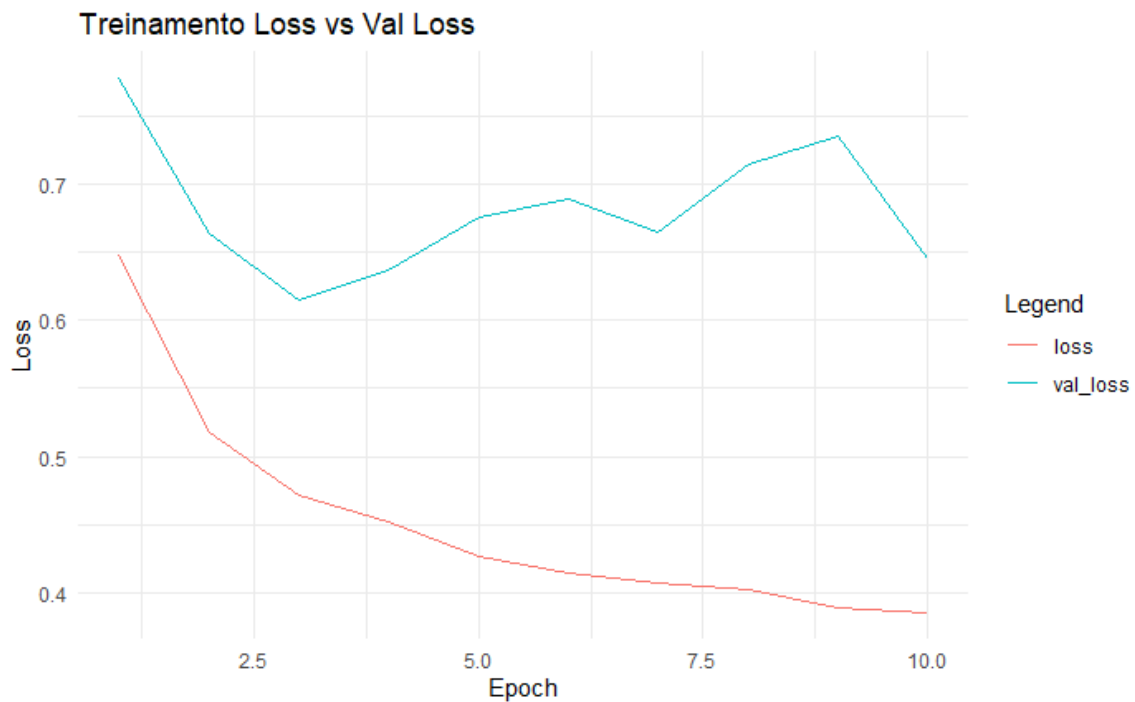


Figura 25: Gráfico das métricas de validação do modelo.

A Figura 26 apresenta as métricas de *accuracy* e *loss* do modelo. A acurácia do treinamento é apresentada em vermelho e a da validação em azul, já o *loss* do treinamento é apresentado em vermelho e da validação em azul, o modelo realiza o treinamento e validação para 10 épocas. A quantidade de *epochs* ou *step* para cada estudo pode variar dependendo do *early stopping* de cada modelo, assim como o *hardware* de cada equipamento e versões dos pacotes usados.



(a) Acurácia do modelo.



(b) Loss do modelo.

Figura 26: Gráfico da acurácia e perda do modelo.

Em contrapartida, ao analisar os estudos dos autores AYAN; ÜNVER (2019); JAIN et al. (2020); KERMANY et al. (2018), o treinamento foi realizado utilizando o modelo VGG sem a diferenciação das classes de pneumonia, as Figuras 27, 28 e 29, representam os resultados para a classificação binária entre sem e com pneumonia nas imagens.

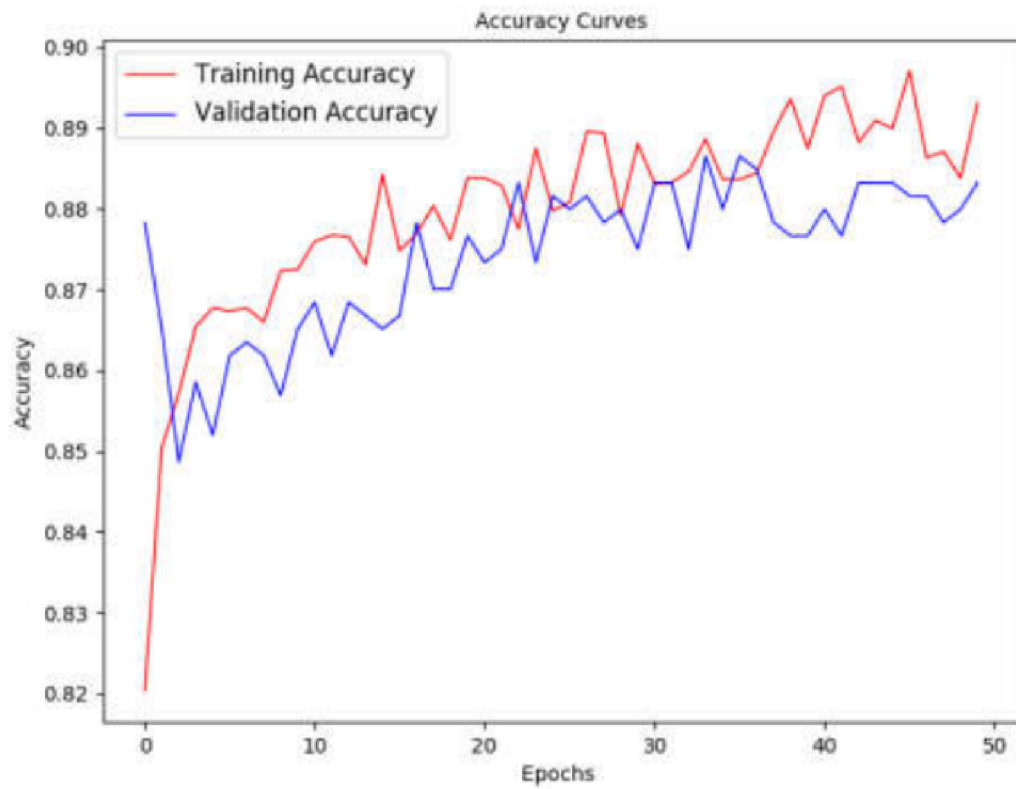


Figura 27: Acurácia do modelo de AYAN; ÜNVER (2019).

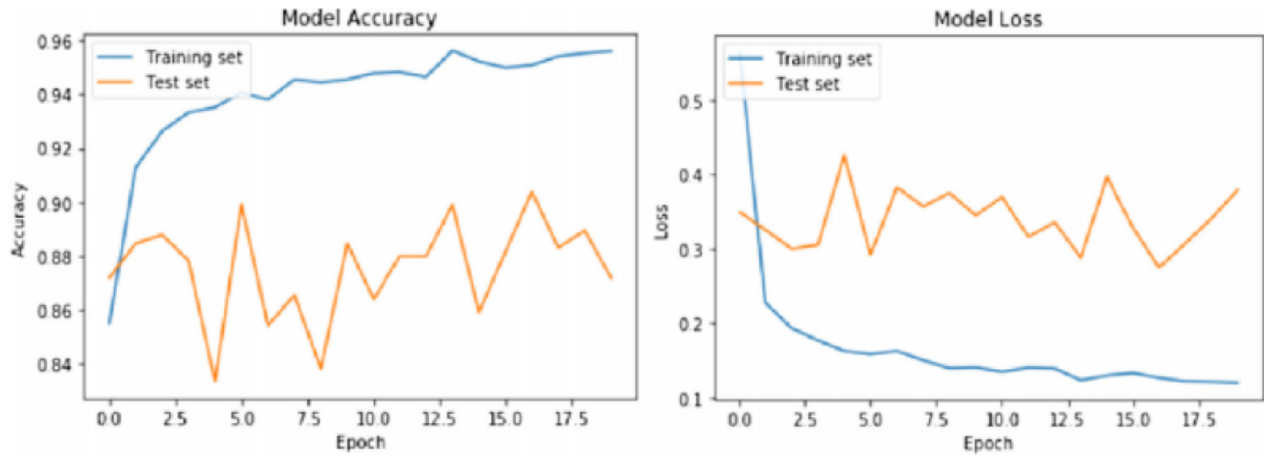


Figura 28: Acurácia do modelo de JAIN et al. (2020).

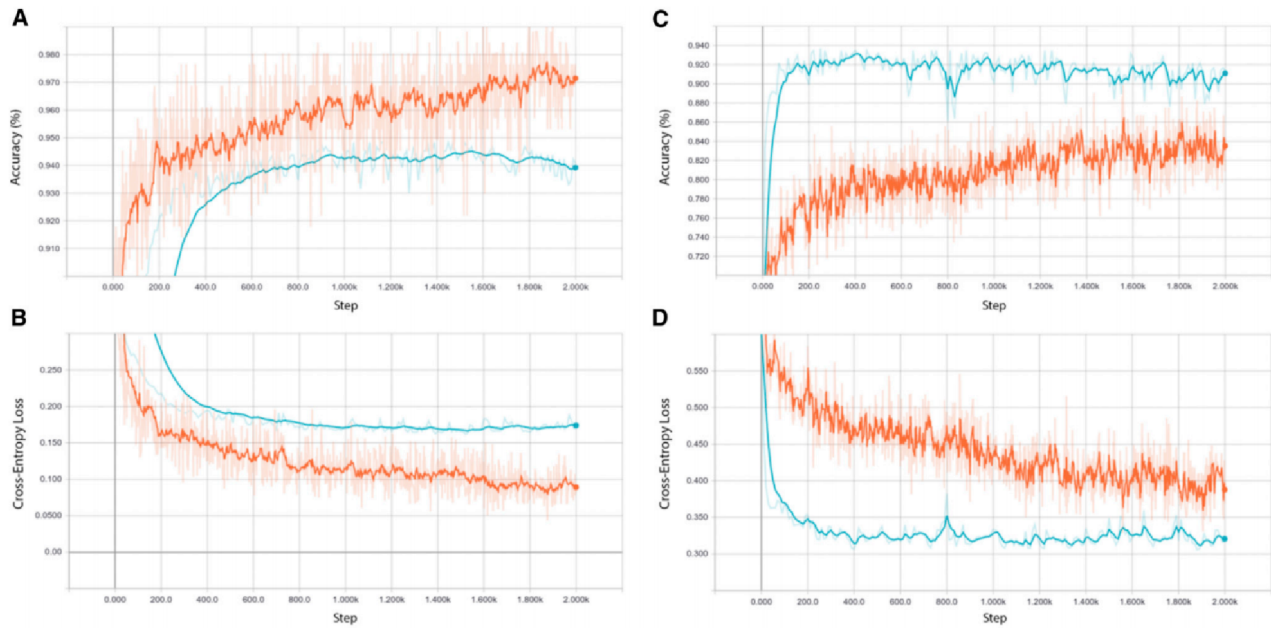


Figura 29: Gráficos descrevem o desempenho do diagnóstico de pneumonia usando imagens de raio-X de tórax nos conjuntos de dados de treinamento e validação usando TensorBoard (A–D) Comparações foram feitas para pneumonia versus normal (A) com perda de entropia cruzada plotada contra a etapa de treinamento (B), bem como comparações entre pneumonia bacteriana e pneumonia viral (C) e a perda de entropia cruzada associada (D) KERMANY et al. (2018).

É possível notar na Tabela 6 que os valores obtidos para o modelo proposto são ligeiramente inferiores em acurácia e perda de aprendizado. Essa diferença nos valores pode ser compreendida analisando a quantidade de classes aprendidas em cada modelo. Os autores da referência classificam seus modelos em duas classes, enquanto o proposto em três.

Tabela 6: Comparação das Métricas de Acurácia e Perda

Referência	Acurácia	Perda
AYAN; ÜNVER (2019)	0,87	0,34
BHANDARY et al. (2020)	0,84	0,11
JAIN et al. (2020)	0,95	0,12
KERMANY et al. (2018)	0,96	0,17
REZAEI; BADIEI; MESHGINI (2020)	0,99	0,02
Modelo Proposto	0,88	0,39

4.3 Matriz de Confusão

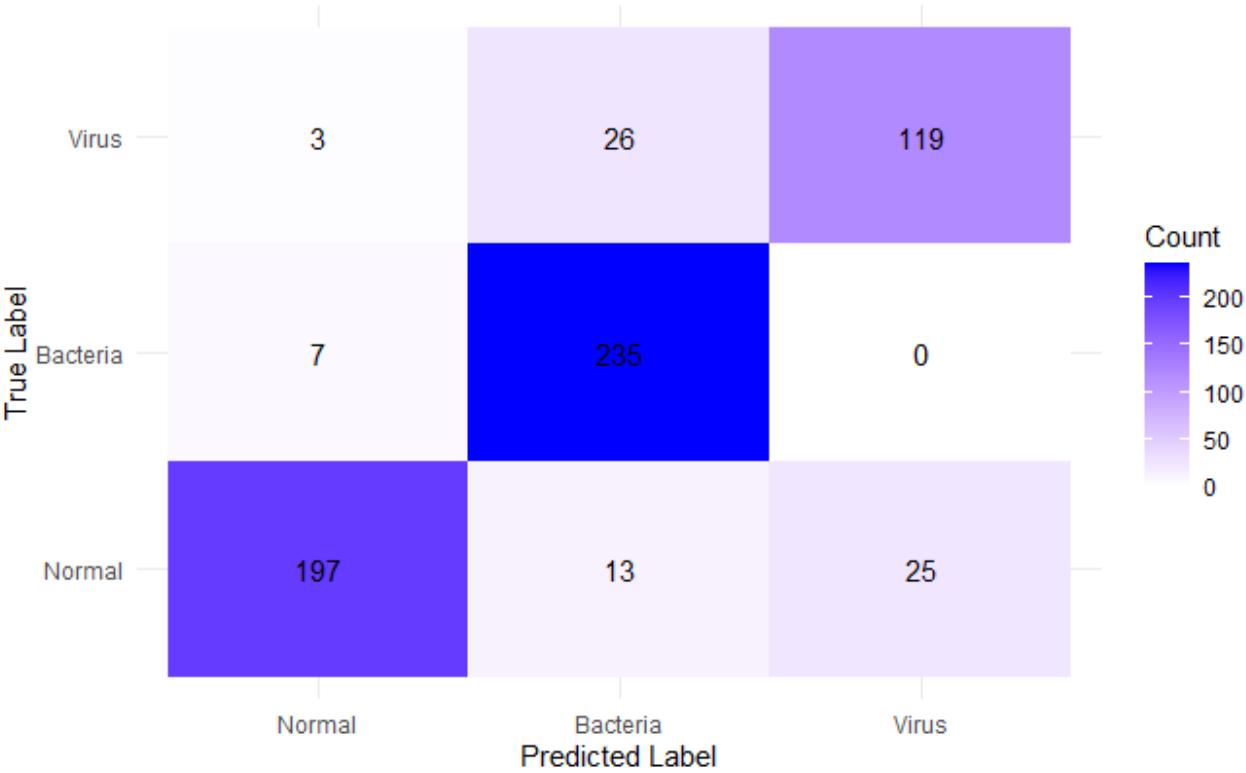
Para testar o modelo, foram utilizadas imagens não anteriormente usadas no treinamento e na validação, essas imagens são inéditas para o sistema. A Tabela 2 apresenta a quantidade de imagens separadas para o modelo.

A matriz de confusão obtida ao realizar o teste do modelo proposto é:

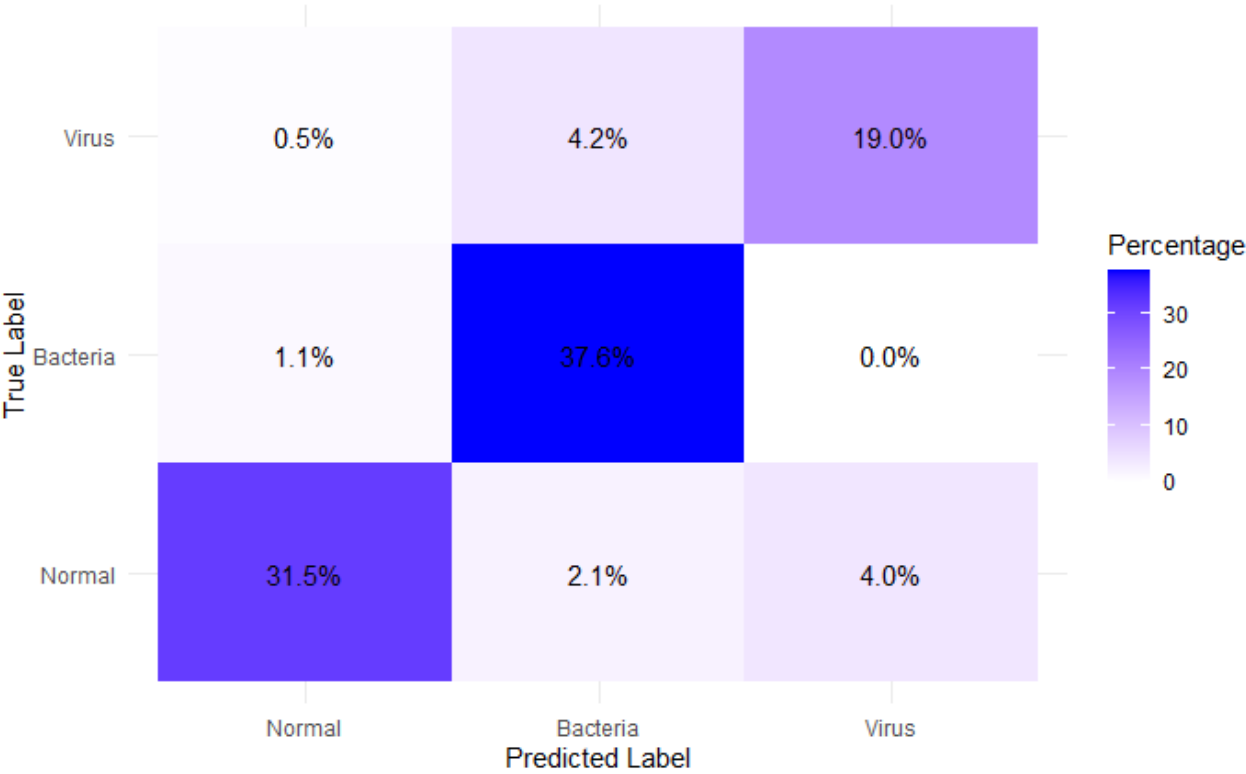
Tabela 7: Matriz de Confusão

Classe	Normal	Bactéria	Vírus
Normal	197 (TP)	13 (FP)	25 (FP)
Bactéria	7 (FN)	235 (TP)	0 (FP)
Vírus	3 (FN)	26 (FN)	119 (TP)

A partir da matriz de confusão apresentada na Figura 30, que apresenta no lado esquerdo a *True Label* que corresponde à nossa classe verdadeira aprendida no treinamento da rede neural, enquanto abaixo temos *Predicted Label* que corresponde à classe predita quando a rede analisa a imagem e a classifica a partir da sua análise. A Figura 30 (a) apresenta a matriz de confusão em contagens e a 30 (b) em valores de porcentagens.



(a) Matriz de Confusão Contagem.



(b) Matriz de Confusão Normalizada Porcentagem.

Figura 30: Matriz de confusão para contagem de acerto no teste do treinamento.

Analizando as matrizes de confusão das figuras 31, 32 e 33, notamos que a maioria dos autores de referência como IBRAHIM et al. (2021); AYAN; ÜNVER (2019) possui o foco de suas análises em classificação entre as classes normal e pneumonia, como em diversos modelos de redes neurais, e outros como TSIKNAKIS et al. (2020) fazem sua análise na classificação de normal, pneumonia bacteriana, pneumonia viral e Covid-19.

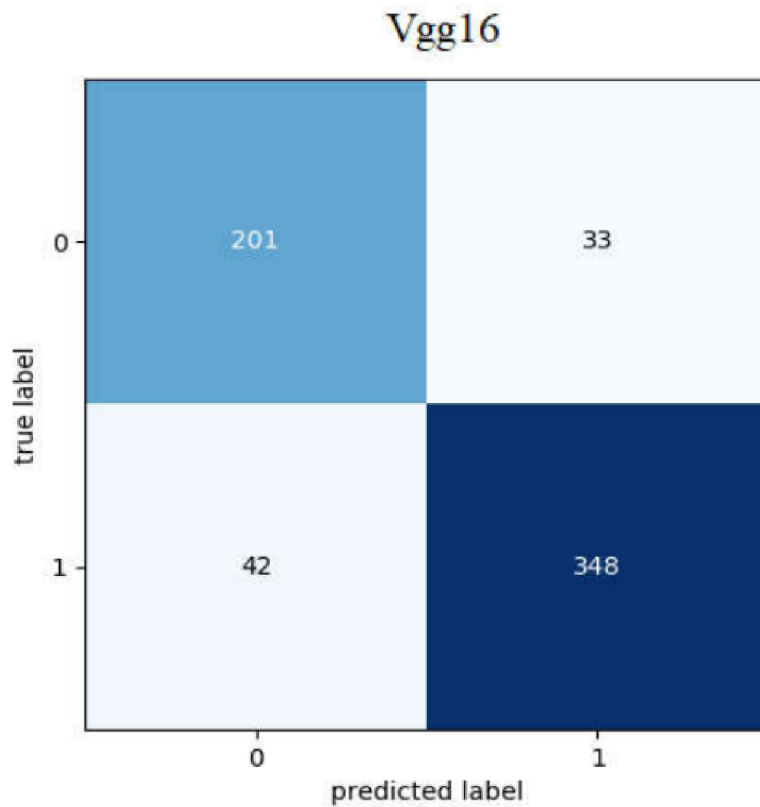


Figura 31: Matriz de Confusão de AYAN; ÜNVER (2019).

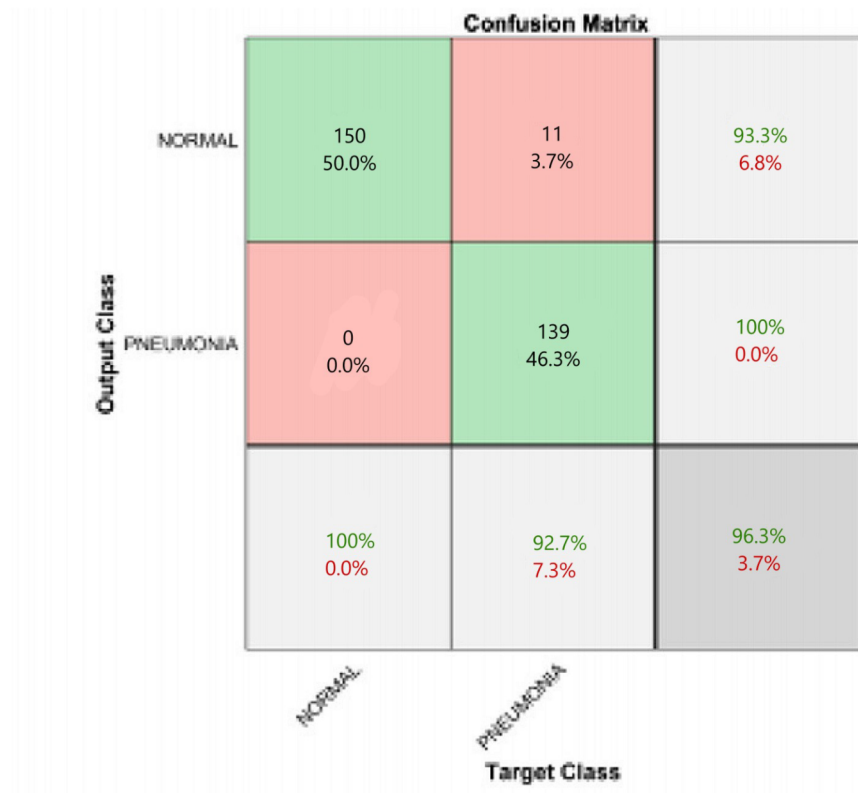


Figura 32: Matriz de Confusão de IBRAHIM et al. (2021).

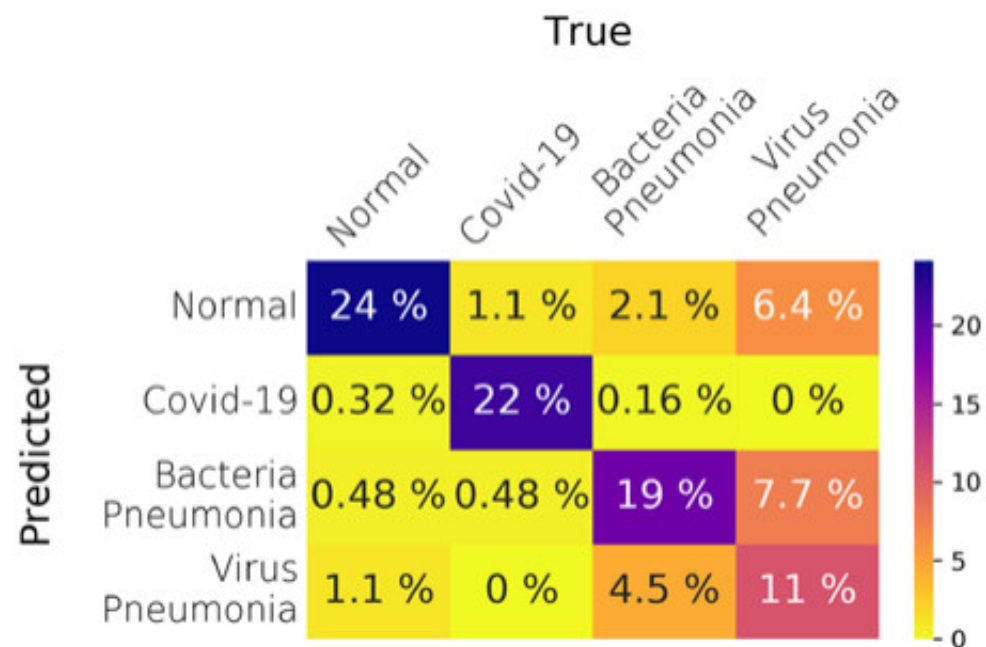


Figura 33: Matriz de Confusão de TSIKNAKIS et al. (2020).

Usando a Tabela 2 o teste do modelo utilizou 625 imagens, sendo 235 normal, 242 bacterianas e 148 para vírus. Foi gerada uma matriz de confusão normalizada para cada caso assim verificando a acurácia com os totais de cada classe.

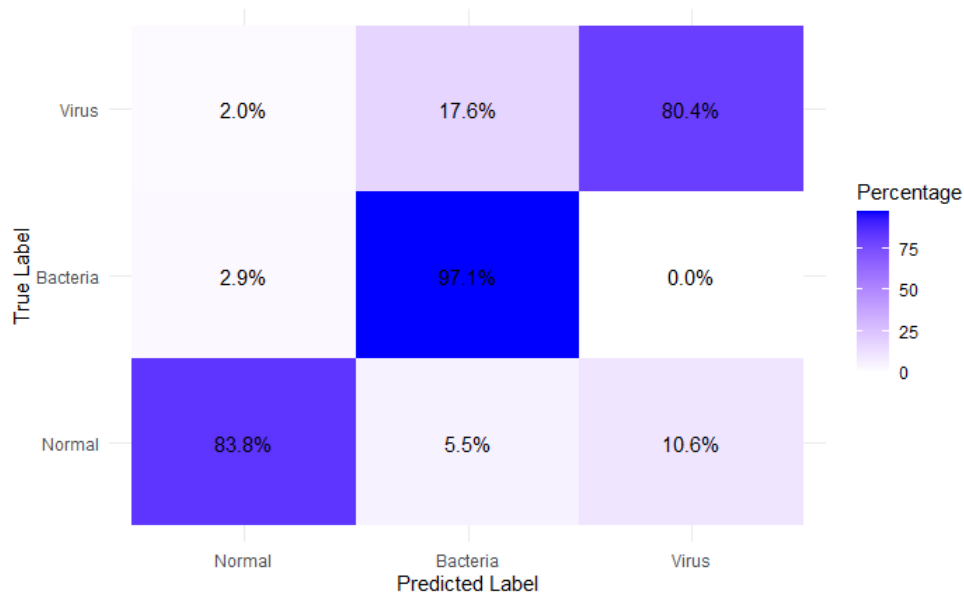


Figura 34: Matriz de confusão normalizada em porcentagem.

Como observado na Figura 34, a classe normal que corresponde aos casos sem pneumonia, com o total de 235 imagens, o modelo previu 83.8% do total das imagens da classe verdadeira como verdadeiros positivos, 5.5% como falsos negativos para pneumonia bacteriana e 10.6% como falso positivo para pneumonia viral. Para a classe de pneumonia bacteriana, com o total de 242 imagens, o modelo previu 97.1% do total como verdadeiro positivo, 2.9% para falso negativo para a classe normal e 0.0% para a classe de pneumonia viral. A classe pneumonia viral possui 148 imagens, o modelo previu 80.4% do total das imagens para a classe de verdadeiro positivo de pneumonia viral, 17.6% de falso positivo para a classe de pneumonia bacteriana e 2.0% de falso negativo para a classe normal.

4.4 Análise das Métricas: Precisão, Recall e F1-score

A partir da matriz de confusão da Figura 30 (a) podemos realizar o cálculo das métricas de precisão *recall* e F1-score.

Classe Normal

- TP: 197
- FP: 7 + 3 = 10 (Predições incorretas para Normal)
- FN: 13 + 25 = 38 (Predições perdidas para Normal)

$$\text{Precisão} = \frac{197}{197 + 10} \approx 0.951 \quad (4.1)$$

$$\text{Recall} = \frac{197}{197 + 39} \approx 0.835 \quad (4.2)$$

$$F1 = 2 \times \frac{0.951 \times 0.835}{0.951 + 0.835} \approx 0.888 \quad (4.3)$$

Classe Bactéria

- TP: 235
- FP: $13 + 25 = 38$ (Predições incorretas para bactéria)
- FN: $7 + 0 = 7$ (Predições perdidas para bactéria)

$$\text{Precisão} = \frac{235}{235 + 39} \approx 0.858 \quad (4.4)$$

$$\text{Recall} = \frac{235}{235 + 7} \approx 0.971 \quad (4.5)$$

$$F1 = 2 \times \frac{0.858 \times 0.971}{0.858 + 0.971} \approx 0.911 \quad (4.6)$$

Classe Vírus

- TP: 119
- FP: $25 + 0 = 25$ (Predições incorretas para vírus)
- FN: $3 + 26 = 29$ (Predições perdidas para vírus)

$$\text{Precisão} = \frac{119}{119 + 25} \approx 0.826 \quad (4.7)$$

$$\text{Recall} = \frac{119}{119 + 29} \approx 0.804 \quad (4.8)$$

$$F1 = 2 \times \frac{0.826 \times 0.804}{0.826 + 0.804} \approx 0.815 \quad (4.9)$$

A tabela 8 apresenta o resultado dos cálculos para as métricas de precisão recall e F1-score para os casos.

Tabela 8: Resultados das Métricas

Casos	Precisão	Recall	F1-Score
Normal	95,1	83,5	88,8
Bactéria	85,8	97,1	91,1
Vírus	82,6	80,4	81,5

A comparação dos resultados com os autores de referência na Tabela 9, podemos notar que as métricas usadas para comparar a acurácia dos modelos proposto, usando o VGG16,

possuem valores próximos, porém os autores nem sempre utilizam as mesmas métricas para a avaliação dos modelos. Pode-se analisar que a métrica de *recall* é usada apenas por dois dos autores de referência.

Tabela 9: Comparação das Métricas no Teste do Modelo

Modelos	Referência	Accuracy	Precisão	Recall	F1-score
VGG16	BHANDARY et al. (2020)	84	82.19	-	84.76
VGG16	IBRAHIM et al. (2021)	93.42	-	-	-
VGG16	JAIN et al. (2020)	95	87.16	96	90
VGG16	KERMANY et al. (2018)	96	-	-	-
VGG16	REZAEI; BADIEI; MESHGINI (2020)	99	99.50	99.20	99
VGG16	Modelo proposto	88,1	89.10	87	88

A Tabela 14 apresenta os valores das métricas para o teste. Notamos que os valores do método proposto são ligeiramente inferiores aos dos demais métodos comparados. Essa diferença pode estar relacionada à maior complexidade envolvida na separação de três classes (Normal, Bactéria e Vírus) em comparação com a maioria dos estudos, que frequentemente abordam apenas duas classes (Normal e Pneumonia). Contudo, os valores calculados na Tabela 8 demonstram que as métricas do modelo proposto permanecem competitivas, sugerindo que ele é robusto mesmo em condições de maior complexidade.

4.5 Curva ROC

A curva ROC é amplamente utilizada para avaliar o desempenho de modelos de classificação, especialmente em problemas relacionados ao diagnóstico médico e à detecção de padrões. Nesta seção, apresentamos a análise da curva ROC e da área sob a curva AUC para o modelo desenvolvido, considerando as diferentes classes do problema abordado. As métricas fundamentais, como a taxa de verdadeiros positivos TPR (sensibilidade) e a taxa de falsos positivos FPR, a especificidade, foram previamente descritas na seção de métodos, permitindo a construção da curva ROC e o cálculo da AUC como medidas de desempenho.

Cálculo das Métricas

A seguir, calculamos as métricas de desempenho para cada classe (Normal, Bactéria e Vírus), com base na matriz de confusão fornecida.

Classe Normal

Para a classe *Normal*, temos os seguintes valores:

- TP = 197
- FN = 10

- $FP = 38$
- $TN = 242 + 148 - (197 + 10 + 38) = 145$
- **TPR (Taxa de Verdadeiro Positivo):**

$$TPR_{\text{Normal}} = \frac{TP}{TP + FN} = \frac{197}{197 + 10} = \frac{197}{207} \approx 0.952$$

- **FPR (Taxa de Falso Positivo):**

$$FPR_{\text{Normal}} = \frac{FP}{FP + TN} = \frac{38}{38 + 145} = \frac{38}{183} \approx 0.208$$

- **Sensibilidade (ou TPR):**

$$\text{Sensibilidade}_{\text{Normal}} = TPR_{\text{Normal}} = 0.952$$

- **Especificidade:**

$$\text{Especificidade}_{\text{Normal}} = \frac{TN}{TN + FP} = \frac{145}{145 + 38} = \frac{145}{183} \approx 0.792$$

Classe Bactéria

Para a classe *Bactéria*, temos os seguintes valores:

- $TP = 235$
- $FN = 33$
- $FP = 13$
- $TN = 235 + 148 - (235 + 33 + 13) = 102$
- **TPR (Taxa de Verdadeiro Positivo):**

$$TPR_{\text{Bactéria}} = \frac{TP}{TP + FN} = \frac{235}{235 + 33} = \frac{235}{268} \approx 0.877$$

- **FPR (Taxa de Falso Positivo):**

$$FPR_{\text{Bactéria}} = \frac{FP}{FP + TN} = \frac{13}{13 + 102} = \frac{13}{115} \approx 0.113$$

- **Sensibilidade (ou TPR):**

$$\text{Sensibilidade}_{\text{Bactéria}} = TPR_{\text{Bactéria}} = 0.877$$

- **Especificidade:**

$$\text{Especificidade}_{\text{Bactéria}} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{102}{102 + 13} = \frac{102}{115} \approx 0.887$$

Classe Vírus

Para a classe *Vírus*, temos os seguintes valores:

- $\text{TP} = 119$
- $\text{FN} = 29$
- $\text{FP} = 25$
- $\text{TN} = 235 + 242 - (119 + 29 + 25) = 303$
- **TPR (Taxa de Verdadeiro Positivo):**

$$\text{TPR}_{\text{Vírus}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{119}{119 + 29} = \frac{119}{148} \approx 0.804$$

- **FPR (Taxa de falso positivo):**

$$\text{FPR}_{\text{Vírus}} = \frac{\text{FP}}{\text{FP} + \text{TN}} = \frac{25}{25 + 303} = \frac{25}{328} \approx 0.076$$

- **Sensibilidade (ou TPR):**

$$\text{Sensibilidade}_{\text{Vírus}} = \text{TPR}_{\text{Vírus}} = 0.804$$

- **Especificidade:**

$$\text{Especificidade}_{\text{Vírus}} = \frac{\text{TN}}{\text{TN} + \text{FP}} = \frac{303}{303 + 25} = \frac{303}{328} \approx 0.924$$

Tabela 10: Métricas de Desempenho das Classes

Classe	TPR	FPR	Sensibilidade	Especificidade
Normal	0,952	0,208	0,952	0,792
Bactéria	0,877	0,113	0,877	0,887
Vírus	0,804	0,076	0,804	0,924

A Tabela 10, resume os resultados da avaliação do desempenho do modelo para as classes Normal, Bactéria e Vírus, incluindo os valores de TPR e FPR. A análise para os valores foi feita da seguinte forma:

Classe Normal:

- O TPR de 0,952 indica que o modelo identificou corretamente 95,2% dos casos pertencentes à classe Normal, sugerindo um bom desempenho na detecção dessa classe.
- O FPR de 0,208 revela que apenas 20,8% dos casos das outras classes foram incorretamente classificados como Normal, indicando uma baixa taxa de falsos positivos.

Classe Bactéria:

- O TPR de 0,877 mostra que o modelo é altamente eficaz em identificar corretamente casos dessa classe, com uma taxa de acerto de 87,7%.
- O FPR de 0,113 é relativamente mais alto em comparação com as outras classes, sugerindo que 11,3% dos casos de outras classes foram incorretamente classificados como Bactéria.

Classe Vírus:

- O TPR de 0,804 indica que o modelo identificou corretamente 80,4% dos casos dessa classe, representando um desempenho levemente inferior em relação às outras classes.
- O FPR de 0,076 mostra que 7,6% dos casos de outras classes foram erroneamente classificados como Vírus, indicando um nível aceitável de erros nesse aspecto.

Os valores indicam que o modelo possui um desempenho robusto, mas com algumas limitações. O alto TPR para Bactéria demonstra excelente capacidade de detecção, enquanto o maior FPR nessa classe sugere a necessidade de ajuste para evitar diagnósticos excessivos. Por outro lado, a classe Vírus apresenta o menor TPR, indicando que melhorias podem ser implementadas para aumentar a sensibilidade. No caso da classe Normal, observa-se um desempenho equilibrado, com bom TPR e baixo FPR, refletindo confiabilidade no diagnóstico de casos normais.

4.6 Cálculo da AUC para o modelo

Consideremos as taxas de verdadeiros positivos (TPR) e falsos positivos (FPR) que já discutimos para cada classe:

Cálculo da AUC:

Uma das maneiras de calcular a AUC, dado que temos pontos TPR e FPR, é usar a regra do trapézio para encontrar a área sob a curva. A AUC foi calculada utilizando a regra do trapézio, que envolve a soma das áreas dos trapézios entre os pontos de dados:

Organizando os Dados

Para calcular a AUC corretamente, precisamos garantir que os pontos FPR, TPR estejam organizados de forma sequencial, ou seja, o FPR não pode diminuir de um ponto para o próximo, pois isso distorceria o cálculo da área sob a curva.

A Tabela 11 apresenta como os pontos estão organizados de forma crescente de FPR. A partir dos valores, podemos utilizar a regra dos trapézios formados entre pontos consecutivos de FPR e TPR e calcular a curva ROC.

Tabela 11: Pontos Organizados para Cálculo da AUC

FPR	TPR	
0,000	0,000	
0,208	0,952	(Normal)
0,076	0,804	(Vírus)
0,113	0,877	(Bactéria)
1,000	1,000	

Os pontos dados para FPR e TPR estão contidos na Tabela 11.

Calculamos a área de cada trapézio entre os pontos consecutivos:

Primeiro trapézio Entre $FPR_1 = 0,000$ e $FPR_2 = 0,208$:

$$\frac{(0,208 - 0,000) \times (0,952 + 0,000)}{2} = 0,099088$$

Segundo trapézio Entre $FPR_2 = 0,208$ e $FPR_3 = 0,076$:

$$\frac{(0,076 - 0,208) \times (0,804 + 0,952)}{2} = -0,115896$$

Terceiro trapézio Entre $FPR_3 = 0,076$ e $FPR_4 = 0,113$:

$$\frac{(0,113 - 0,076) \times (0,877 + 0,804)}{2} = 0,031068$$

Quarto trapézio Entre $FPR_4 = 0,113$ e $FPR_5 = 1,000$:

$$\frac{(1,000 - 0,113) \times (1,000 + 0,877)}{2} = 0,832445$$

Soma das Áreas

Somando as áreas dos trapézios:

$$AUC = 0,099088 + (-0,115896) + 0,031068 + 0,832445 = 0,8467$$

A AUC resultante de aproximadamente 0,8467 indica que o modelo possui uma boa capacidade de discriminar entre as classes positiva e negativa. Quanto mais próxima de 1 for a AUC, melhor será o desempenho do modelo.

4.7 Cálculo da Curva ROC e AUC para Múltiplas Classes

Para calcular a AUC para cada classe de um modelo com múltiplas classes, o procedimento básico segue o método de *One-vs-Rest* em tradução direta (um contra todos). Para cada classe, o cálculo da curva ROC e da AUC, tratando essa classe como positiva e todas as outras como negativas, é o processo de cálculo para cada classe individualmente, com base na sua matriz de confusão da Tabela 7.

Organização dos Pontos de FPR e TPR

A Tabela 10 apresenta os valores de TPR e FPR para as classes Normal, Bactéria e Vírus. Para cada classe, consideramos os pontos (0, 0) e (1, 1) no início e no final da curva, respectivamente.

4.8 Cálculo da AUC para cada classe (One-vs-Rest)

Para calcular a AUC (Área sob a Curva ROC) para cada classe utilizando a estratégia One-vs-Rest, seguimos os passos descritos abaixo.

4.9 Classe Normal (One-vs-Rest)

Considerando a classe **Normal** como positiva, e as classes **Bacteria** e **Virus** como negativas, os pontos da curva ROC são os seguintes:

- FPR (Taxa de Falsos Positivos) = 0,208
- TPR (Taxa de Verdadeiros Positivos) = 0,952

Usamos a fórmula da AUC com a regra dos trapézios para calcular a AUC para a classe **Normal**:

$$AUC_{\text{Normal}} = \frac{(1.000 - 0.208) \times (1.000 + 0.952)}{2} = \frac{0.792 \times 1.952}{2} = 0.774$$

4.10 Classe Bactéria (One-vs-Rest)

Considerando a classe **Bacteria** como positiva, e as classes **Normal** e **Virus** como negativas, os pontos da curva ROC são os seguintes:

- FPR (Taxa de Falsos Positivos) = 0,076
- TPR (Taxa de Verdadeiros Positivos) = 0,804

A AUC para a classe **Bacteria** é calculada da mesma forma:

$$AUC_{\text{Bacteria}} = \frac{(1.000 - 0.076) \times (1.000 + 0.804)}{2} = \frac{0.924 \times 1.804}{2} = 0.834$$

4.11 Classe Vírus (One-vs-Rest)

Considerando a classe **Virus** como positiva, e as classes **Normal** e **Bacteria** como negativas, os pontos da curva ROC são os seguintes:

- FPR (Taxa de Falsos Positivos) = 0,113

- TPR (Taxa de Verdadeiros Positivos) = 0,877

A AUC para a classe **Virus** é:

$$AUC_{\text{Virus}} = \frac{(1.000 - 0.113) \times (1.000 + 0.877)}{2} = \frac{0.887 \times 1.877}{2} = 0.834$$

4. AUC Média

A AUC Média é calculada pela média das AUC de cada classe:

$$AUC \text{ Média} = \frac{AUC \text{ Normal} + AUC \text{ Bacteria} + AUC \text{ Virus}}{3} = 0.814$$

Resumo dos Resultados

Tabela 12: Resultados da AUC para Cada Classe (One-vs-Rest)

Classe	AUC
Normal	0.774
Bactéria	0.834
Vírus	0.834
Média Geral da AUC	0.814

4.12 Análise de Normal vs Pneumonia

Reorganização da Matriz de Confusão

A matriz de confusão é reorganizada considerando apenas duas classes: **Normal** como positiva e **Pneumonia** (Bactéria + Vírus) como negativa.

Tabela 13: Matriz de confusão para Normal vs Pneumonia.

Real / Predito	Normal	Pneumonia	Total
Normal	197	38	235
Pneumonia	10	373	383
Total	207	411	618

Cálculo do TPR e FPR

O **TPR** é a proporção de positivos reais (**Normal**) corretamente identificados.

$$TPR = \frac{197}{197 + 38} = \frac{197}{235} \approx 0.839 \quad (4.10)$$

Temos $FP = 38$ e $TN = 197$ então, o cálculo da **FPR** é:

$$FPR = \frac{38}{38 + 197} = \frac{38}{235} \approx 0.1617 \quad (4.11)$$

Portanto, a **FPR** para o caso Normal vs Pneumonia é aproximadamente 0.1617 ou 16.17%.

A Área Sob a Curva ROC (AUC) é calculada utilizando a regra dos trapézios, que estima a área sob a curva ROC com base na média dos valores de TPR (Taxa de Verdadeiros Positivos) entre dois pontos consecutivos, multiplicada pela diferença entre os valores de FPR (Taxa de Falsos Positivos) desses pontos consecutivos.

A fórmula para calcular a **AUC** é:

$$AUC = \sum_{i=1}^{n-1} \frac{(TPR_i + TPR_{i+1})}{2} \times (FPR_{i+1} - FPR_i)$$

Considerando os seguintes pontos da curva ROC:

- Ponto 1: FPR = 0, TPR = 0
- Ponto 2: FPR = 0.1617, TPR = 0.839
- Ponto 3: FPR = 1.0, TPR = 1.0

Agora, aplicando a fórmula:

$$AUC = \frac{(0 + 0.839)}{2} \times (0.1617 - 0) + \frac{(0.839 + 1.0)}{2} \times (1.0 - 0.1617)$$

Calculando as duas áreas do trapézio:

$$AUC = 0.4195 \times 0.1617 + 0.9195 \times 0.8383$$

$$AUC = 0.068 + 0.7712 = 0.8392$$

Portanto, a **AUC** é aproximadamente 0.839

Tabela 14: Comparação das métricas no Teste do modelo

Referência	TPR	FPR
BHANDARY et al. (2020)	0,875	0,0490
IBRAHIM et al. (2021)	0,892	0,0108
JAIN et al. (2020)	-	-
KERMANY et al. (2018)	0,960	0,099
REZAEE; BADIEI; MESHGINI (2020)	0,990	0,005
Modelo proposto para normal vs pneumonia	0,839	0,162
Modelo proposto para normal	0,952	0,208
Modelo proposto para bacteria	0,877	0,113
Modelo proposto para virus	0,804	0,076

A partir dos valores de TPR e FPR, realizamos a plotagem gráfica dos valores representados. A Figura 35 ilustra a curva ROC para a classificação multiclasse, demonstrando a capacidade do modelo em distinguir entre as classes positivas e negativas.

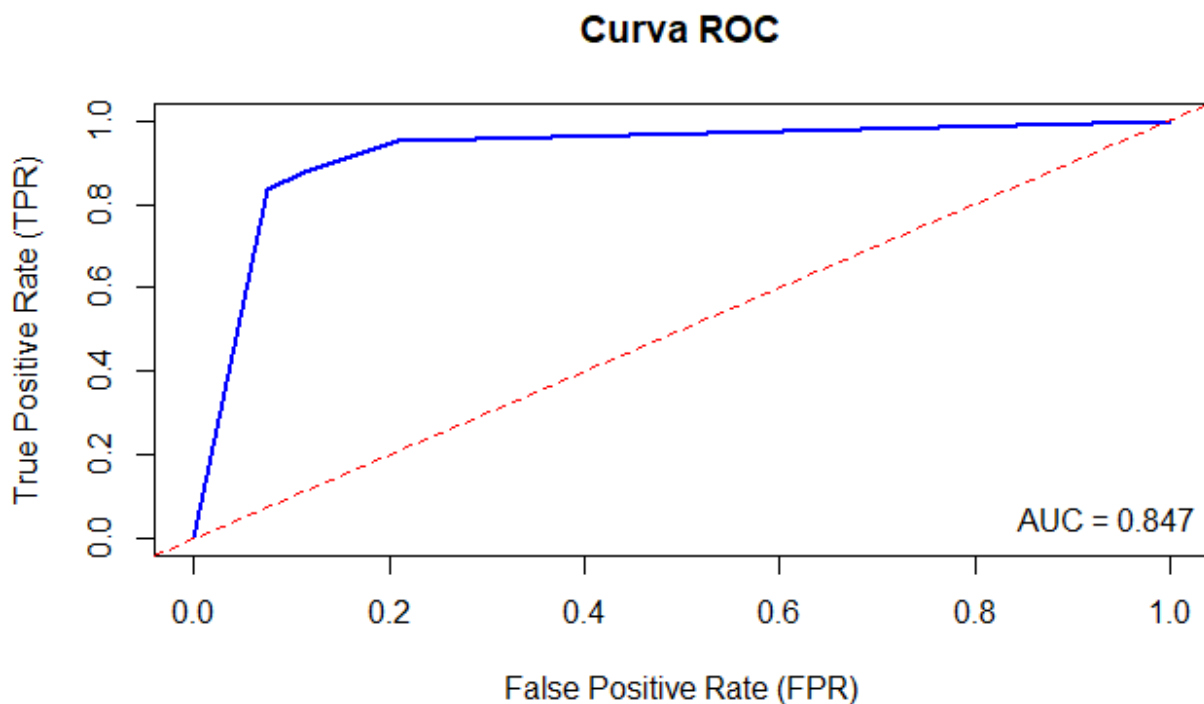


Figura 35: Curva ROC para Classificação multiclasse

A Figura 36 apresenta a curva ROC para cada uma das classes usando o método *One vs rest*, em que cada classe apresenta a comparação de TPR e FPR.

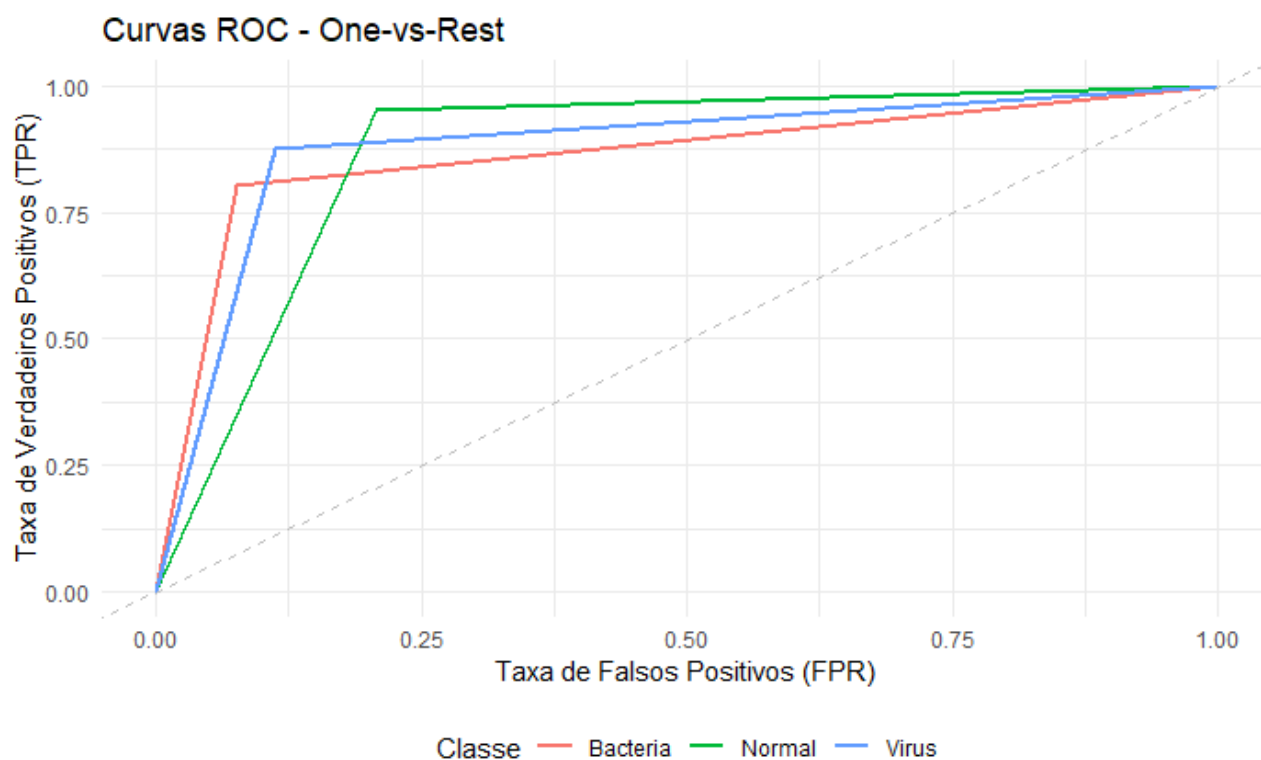


Figura 36: Curva ROC para as classes pelo metodo *one vs rest*

A Figura 37 apresenta a curva ROC para normal vs pneumonia, mostrando a variação entre TPR e FPR.

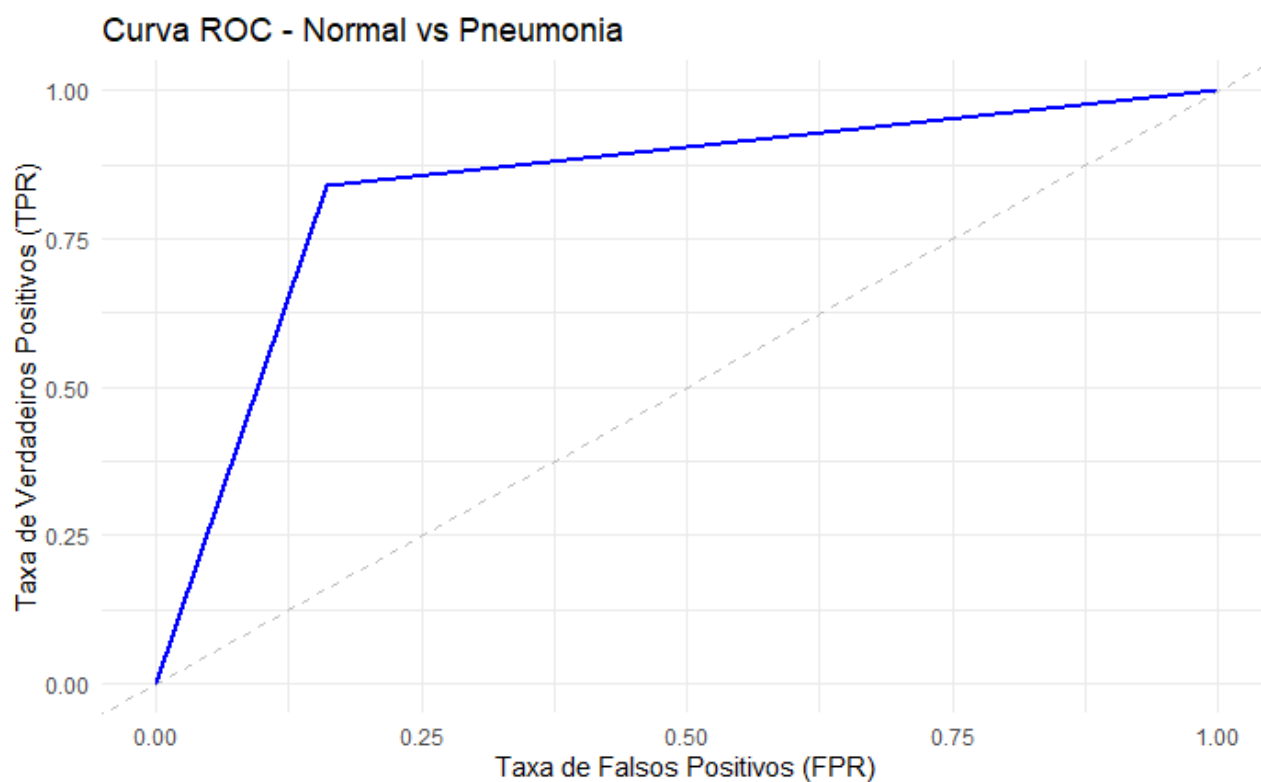


Figura 37: Curva ROC para normal vs pneumonia

A Curva ROC é uma representação gráfica das taxas TPR (Sensibilidade) e FPR para vários limiares de decisão. A curva exibe o desempenho do modelo em distinguir entre as classes positivas e negativas em todos os possíveis pontos de corte. Quanto mais perto da origem (0,0) a curva se mantiver e mais próxima da linha diagonal (linha de chance aleatória) ela for, pior será o desempenho do modelo.

Em comparação aos autores BHANDARY et al. (2020); KERMANY et al. (2018) analisando suas curvas ROC, podemos notar que a distribuição do TPR para o teste dos modelos considera a sensibilidade do modelo em prever as classes corretas do banco de dados e método proposto. Porém, os autores classificam os modelos por normal vs pneumonia, normal, anormal e cancerígeno, e a discussão de seus resultados aborda uma classificação de duas, porém não discutem a separação dos casos de pneumonia bacteriana e viral.

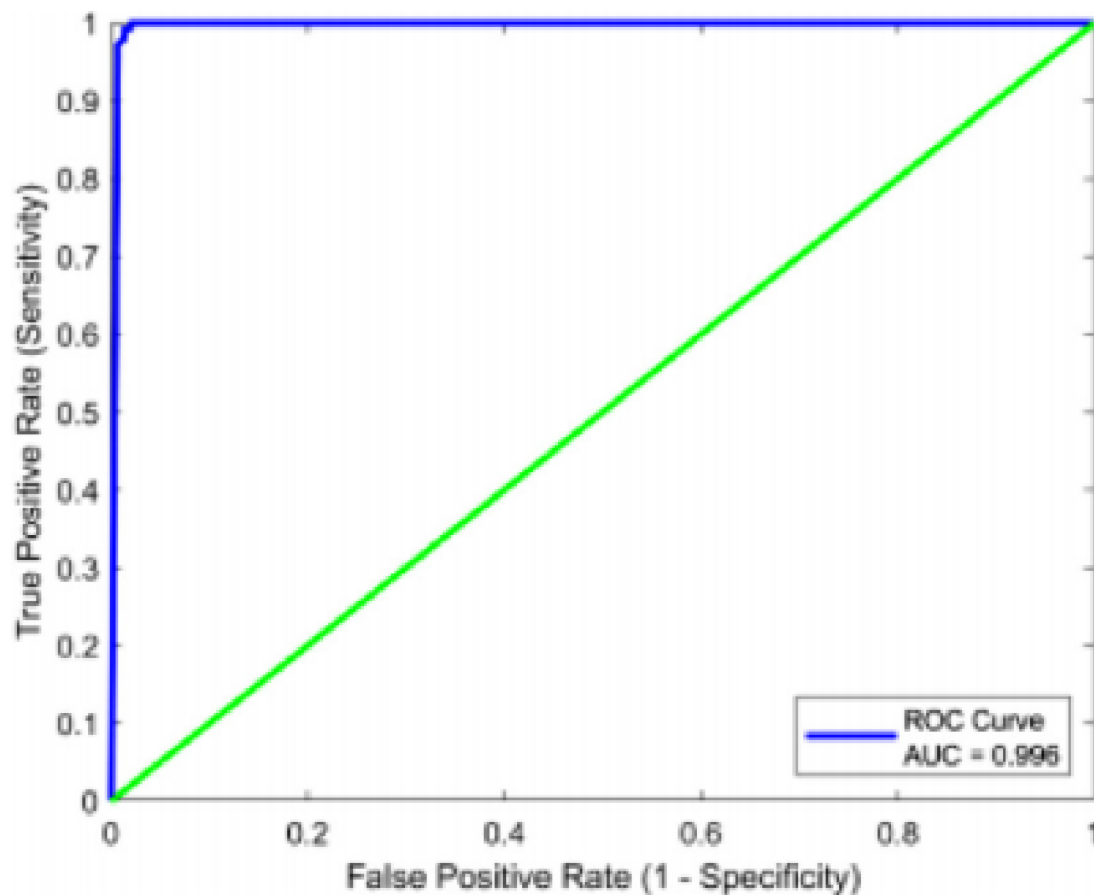


Figura 38: Curva ROC de BHANDARY et al. (2020)

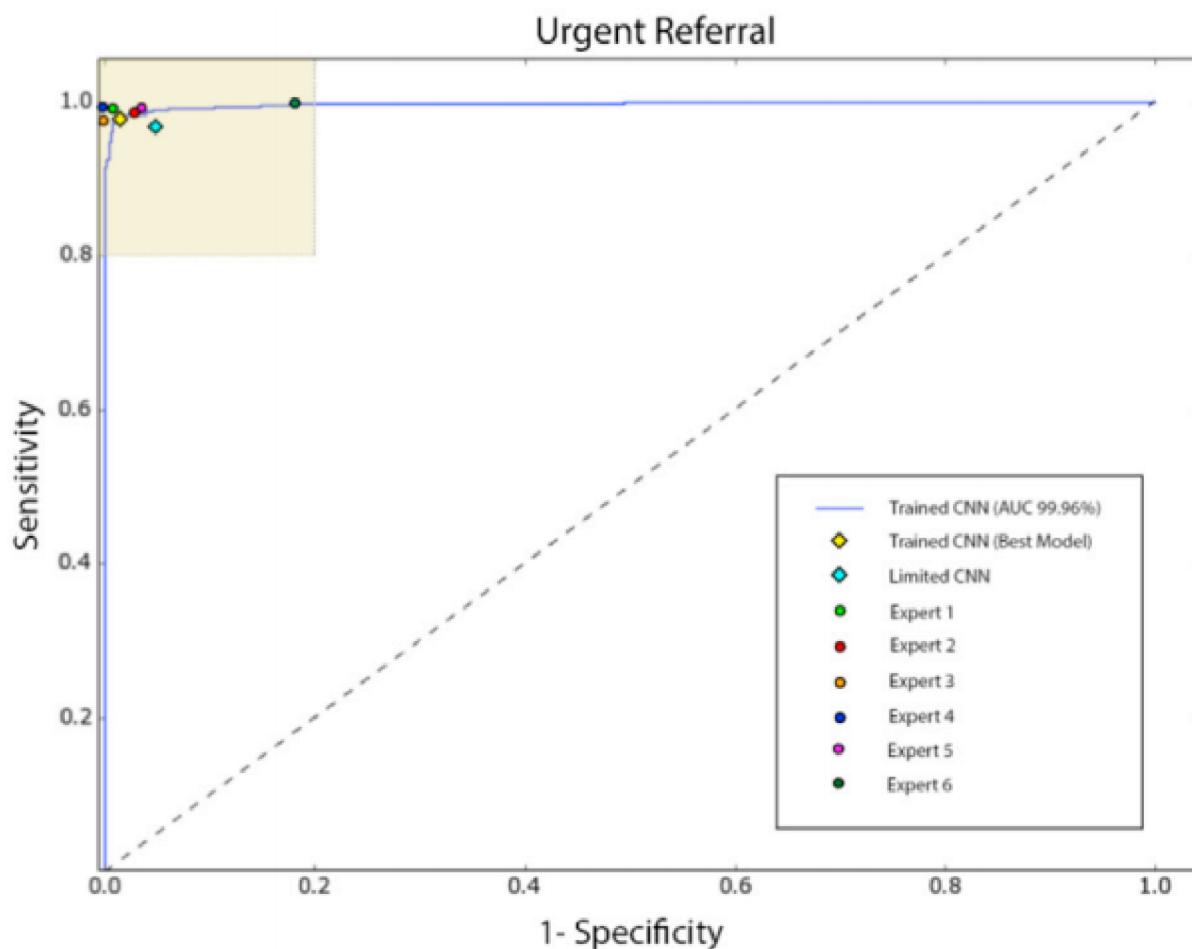


Figura 39: Curva ROC de KERMANY et al. (2018)

A abordagem de amostragem dos resultados do treinamento do modelo considerou a diferenciação dos tipos de pneumonia, utilizando a classificação multiclasse, conforme apresentado na Tabela 15, em que os valores de AUC, mesmo estando inferiores aos demais modelos apresentam valores entre 0,8 a 0,9 o que demonstra que o modelo apresenta um bom desempenho em termos de acertos na classificação das TPR. Isso indica que, embora haja espaço para melhorias, o modelo já realiza uma tarefa diferenciada em relação aos outros, abrindo possibilidades para aprimoramentos futuros.

Tabela 15: Comparação dos Valores de AUC em Modelos de Classificação de Pneumonia

Autor	Dataset	Modelo	AUC (%)
BHANDARY et al. (2020)	Pneumonia	AlexNet	99,6
KERMANY et al. (2018)	Pneumonia vs. Normal	ResNet-50	96,8
KERMANY et al. (2018)	Pneumonia Bacteriana e Viral	ResNet-50	94,0
Este Trabalho	Pneumonia vs. Normal	Modelo Proposto	83,9
Este Trabalho	Sem Pneumonia	Modelo Proposto	77,4
Este Trabalho	Pneumonia Bacteriana	Modelo Proposto	87,7
Este Trabalho	Pneumonia Viral	Modelo Proposto	80,4

A Tabela 16 apresenta a comparação dos valores de sensibilidade e especificidade dos modelos de referência e o modelo proposto. Podemos notar que os valores de especificidade apresentados estão diretamente relacionados com as taxas de aprendizado em distinguir os casos de verdadeiros negativos, os dados apresentam valores abaixo dos demais autores, isso acontece porque o modelo proposto treinou a classificação multiclasse em todos os estágios, a análise dos resultados teve como abordagem sintetizar o resultado como normal vs pneumonia para analisar como os demais autores, essa abordagem faz com que os valores reais de aprendizado sejam reduzidos.

Os valores para sem pneumonia, pneumonia bactéria e pneumonia vírus apresentam os valores reais de aprendizagem do modelo em diferenciação multiclasse das taxas de verdadeiros negativos. Com isso, observamos que os valores obtidos para as classes separadamente apresentam um bom desempenho.

Com isso, observa-se que os valores obtidos para as classes separadas indicam um bom desempenho. Os valores de sensibilidade estão próximos aos de um modelo de classificação binária, o que demonstra que, mesmo com uma maior complexidade de classificação, o sistema obteve uma boa taxa de aprendizado, comparada aos sistemas mais refinados.

Tabela 16: Comparação dos Valores de Sensibilidade e Especificidade em Modelos de Classificação de Pneumonia

Autor	Dataset	Sensibilidade (%)	Especificidade (%)
BHANDARY et al. (2020)	Pneumonia	98,09	95,63
KERMANY et al. (2018)	Pneumonia vs normal	93,02	90,1
KERMANY et al. (2018)	Pneumonia bactéria vs vírus	88,6	90,9
Este Trabalho	Pneumonia vs normal	83,0	83,8
Este Trabalho	Sem Pneumonia	83,9	79,2
Este Trabalho	Pneumonia bactéria	87,6	88,7
Este Trabalho	Pneumonia vírus	80,4	92,4

No âmbito hospitalar, a pneumonia pode ser contraída de várias formas, sendo elas virais e bacterianas. No diagnóstico por imagens, fazer a diferenciação em apenas saudável ou doente

pode acarretar a um tratamento errado e a patologia pode evoluir para uma multirresistência ao tratamento. A maior complexidade do modelo proposto para combater esses problemas é lidar com a classificação multiclasse, que em comparação aos de referência tratam de análises binárias, o que pode resultar em uma redução significativa dos valores de desempenho em comparação com os modelos binários. Isso ocorre porque, ao adicionar mais classes, o modelo precisa aprender a diferenciar entre mais categorias, o que pode afetar a precisão em cada uma das classes individualmente, reduzindo os valores de métricas como AUC, sensibilidade e especificidade. Outro fator importante a ser analisado é o banco de dados, modelos de redes neurais são treinados com bancos de dados massivos, a eficiência de treinamento e aprendizagem das redes são superiores a um modelo treinado com bancos de dados menores.

O modelo VGG é normalmente utilizado para diferenciação binária em imagens de raio-x, o modelo foi utilizado como base e modificado dando origem ao modelo proposto que realiza o treinamento em três classes (normal, pneumonia bacteriana e pneumonia viral), em comparação com a análise dos modelos de referência são realizadas apenas em duas classes (normal e pneumonia),(Covid e bacteria) entre outras formas binárias.

4.13 Auxílio ao diagnostico

Para auxiliar o grande volume de dados do *Dataset* e agilizar o diagnóstico médico o modelo proposto foi desenvolvido para que as saídas do treinamento da rede neural apresente qual a porcentagem de previsão para cada classe. As Figuras 40, 41 e 42 demonstram as saídas após a verificação de predição do *dataset* de teste.

A Figura 40 apresenta a resposta das classe normal, a classe normal corresponde as imagens sem pneumonia, no banco de dados as imagens são rotuladas para fazer o controle das predições e podermos saber quais as imagens que o sistema teve mais dificuldades em diferenciar. podemos notar que na parte superior da imagem aparece a descrição da imagem e na parte inferior as predições do modelo.

O teste gerou a predição pada todas as imagens usadas para o teste do modelo as Figuras 40, 41 e 42 apenas demonstram a saída para que o responsável pelo diagnostico possa ter uma base e agilizar a decisão para a continuação do diagnóstico. O sistema apresenta as predições para facilitar e agilizar, as decisões finais ficam a cargo do responsável medico legal.

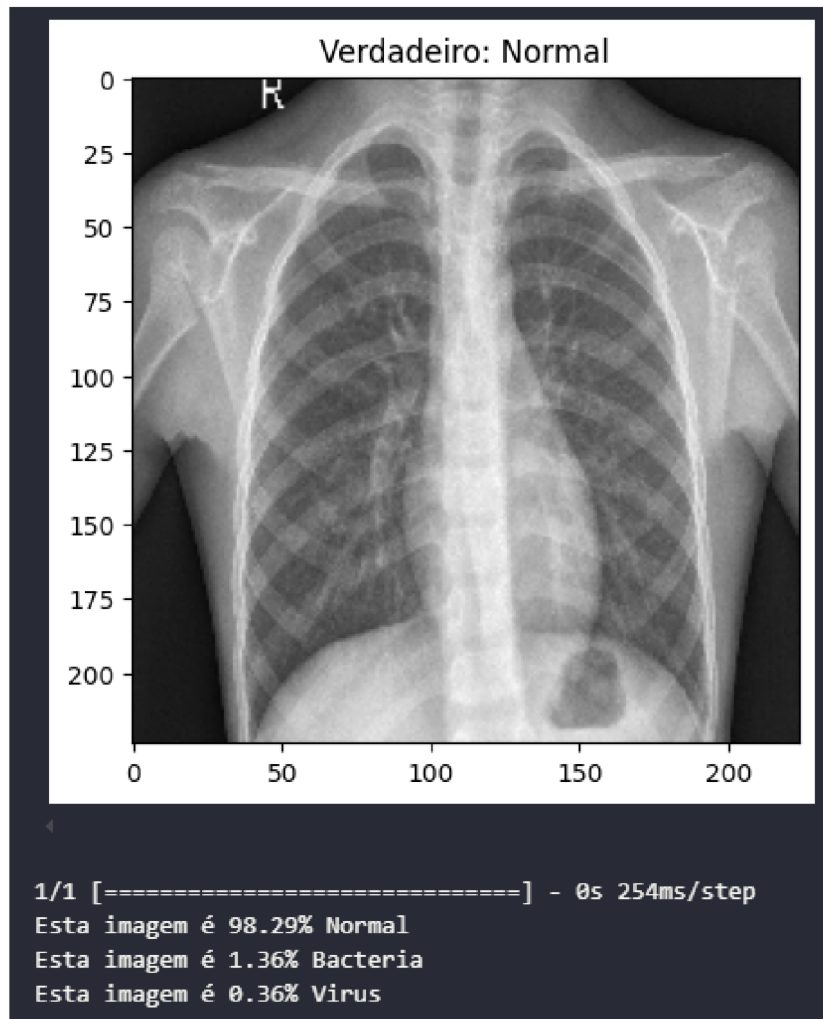


Figura 40: Resposta do modelo para classe normal.

As predições que são expressas pelo modelo, possuem informações relevantes para o diagnóstico médico, as predições são expressas em porcentagem e possuem valores de qual a quantidade reconhecida de cada classes ou padrões de imagem aprendido pela rede neural. Essas predições apresentaram quantidades de normal, bactéria e vírus.

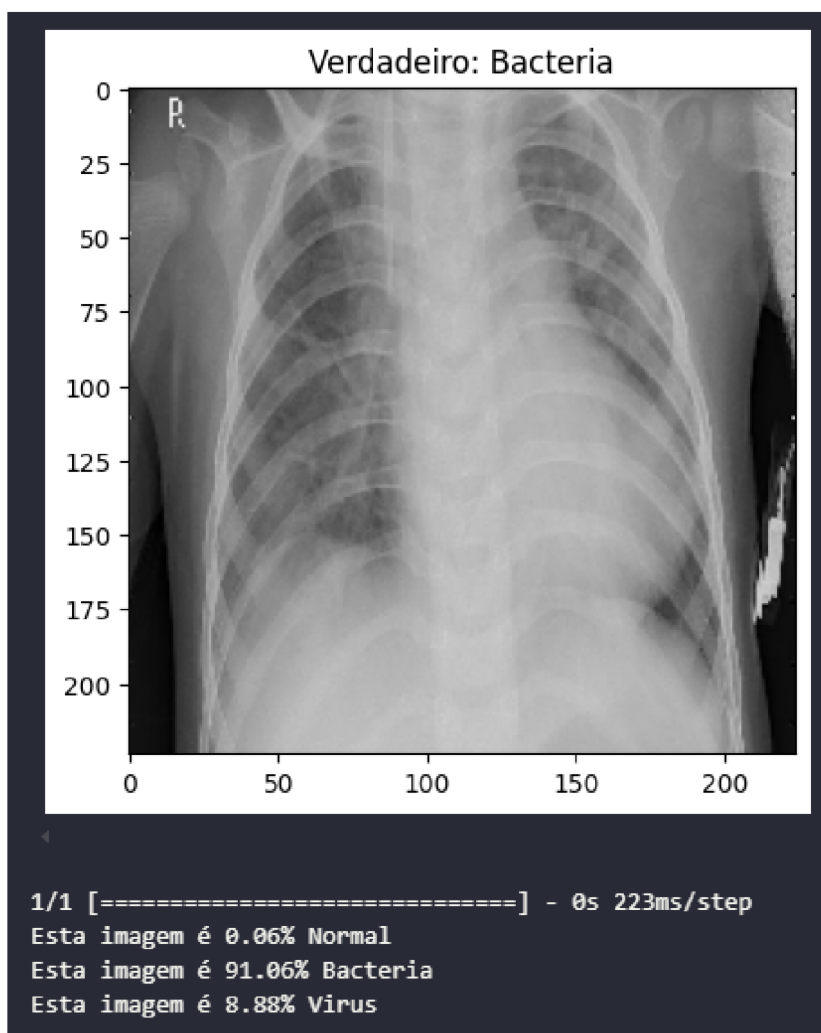


Figura 41: Resposta do modelo para classe bactéria.

Essa visualização é importante por agilizar na tomada de decisão e escolha do tratamento para cada caso, essa possibilidade ajuda a que os verdadeiros casos de pneumonias possam ser tratados mais rápidos, um exemplo seira se em caso hipotético se todas as imagens normais fossem escolhidas aleatoriamente para análise antes das outras, os pacientes que realmente possuísem a pneumonia ficariam esperando o tratamento e as análises para o laudo por muito mais tempo.

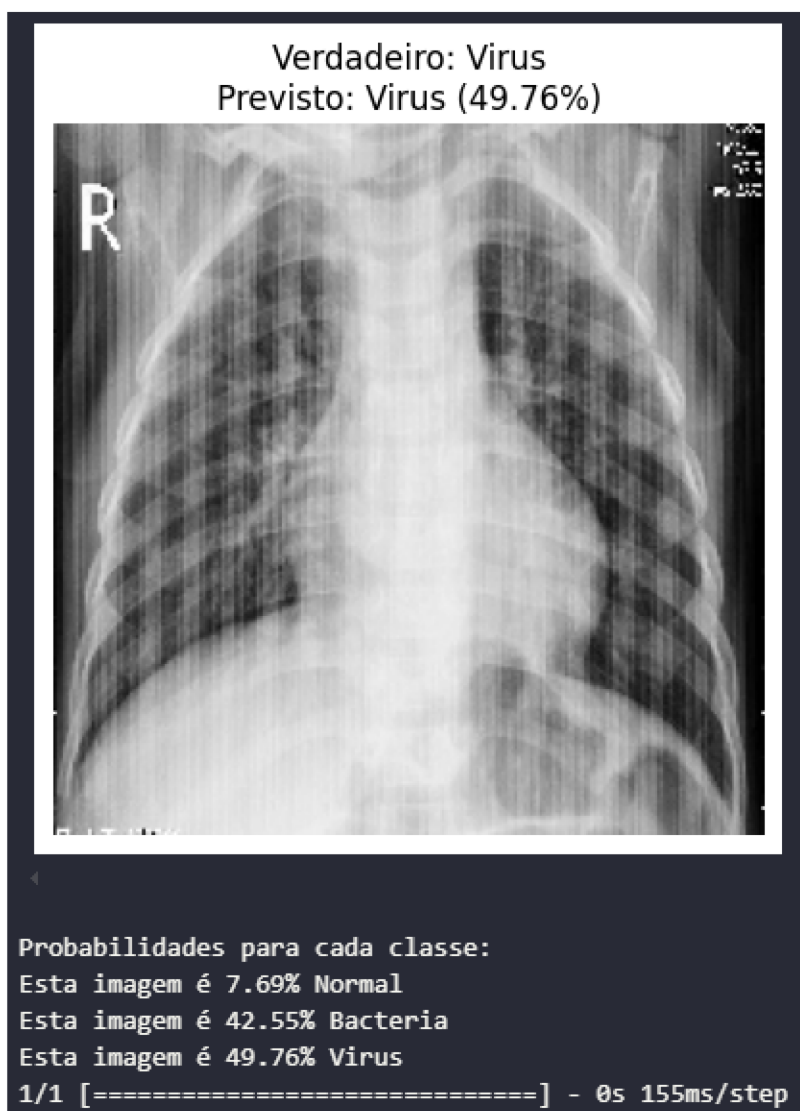


Figura 42: Resposta do modelo para classe vírus.

A escolha de apresentar todas as predições para as imagens é uma ferramenta que ajuda responsáveis pelo diagnóstico, uma visualização mais precisa dos padrões da imagem analisada e uma possível forma de analisarem as imagens e ajudar em suas decisões.

5 CONCLUSÃO

Neste trabalho discutimos a importância da classificação de padrões em imagens médicas como auxílio ao diagnóstico médico. Apresentamos uma classificação multiclasse entre pneumonia bacteriana e viral e sem a presença de pneumonia, a diferenciação das pneumonias para auxiliar os profissionais de saúde ao classificar e separar os tipos de pneumonia, o auxílio de classificar previamente as imagens médicas ajuda a agilizar a escolha de tratamento para cada tipo de pneumonia e seu tempo de espera no diagnóstico. O modelo consiste em uma rede neural convolucional CNN pré-treinada com o modelo base VGG16, complementada com uma camada de *flatten* e duas camadas densas com ativação ReLU e uma camada de saída utiliza ativação *softmax* para a classificação multiclasse.

1. Os resultados obtidos do teste de desempenho geral do modelo apresentaram uma acurácia de 88%, com uma perda de 0,39%. O modelo proposto alcançou uma precisão de 83,5% e uma *recall* de 80,4% para a classe de pneumonia bacteriana, e uma precisão de 80,2% e uma *recall* de 76,3% para a classe de pneumonia viral. Isso sugere que o modelo é capaz de identificar os casos de pneumonia bacteriana e viral, mas pode ter dificuldade em distinguir entre as duas classes.
2. No teste de predição do modelo na Tabela 2, o modelo proposto alcançou uma acurácia de 83,8% para a classe sem pneumonia, e uma acurácia de 97,1% para a classe de pneumonia bacteriana e 80,4% para pneumonia viral. As métricas derivadas da matriz apresentaram uma precisão de 95,1% para a classe normal, 85,8% para a classe bacteriana e 82,6% para a classe viral. Além disso, o modelo alcançou um *recall* de 83,5% para a classe normal, 97,1% para a classe bacteriana e 80,4% para a classe viral. Os valores de *F1-score* também foram calculados e mostram que o modelo alcançou um *F1-score* de 88,8% para a classe normal, 91,1% para a classe bacteriana e 81,5% para a classe viral.
3. A análise das curvas ROC e dos valores de AUC, que avaliaram o desempenho do modelo em classificar corretamente imagens normais e de pneumonia, revelou uma AUC de 83,9%, demonstrando um desempenho competitivo. Em particular, a especificidade (83,8%) e a sensibilidade (83,0%) atingidas pelo modelo indicam um equilíbrio adequado entre a capacidade de identificar corretamente tanto os casos positivos quanto os negativos de pneumonia. O uso das métricas específicas para cada classe (One-vs-Rest) permitiu uma avaliação detalhada de desempenho do modelo proposto, apresentando uma AUC de 77,4% para a classificação de sem pneumonia, AUC de 87,7% para pneumonia bacteriana e AUC de 80,4% para pneumonia viral.
4. A sensibilidade e a especificidade são métricas críticas para avaliar o desempenho de modelos de diagnóstico médico. O modelo proposto alcançou uma sensibilidade de 87,6% e uma especificidade de 88,7% para a classificação de pneumonia bacteriana, para a

classificação de pneumonia viral, o modelo proposto alcançou uma sensibilidade de 80,4% e uma especificidade de 92,4%.

A diferenciação multiclasse utilizada no trabalho para o treinamento, validação e teste é importante por ser mais adequada para aplicações práticas em diagnósticos médicos assistidos por computador. Essa abordagem é particularmente importante para a aplicabilidade clínica, pois a identificação da precisão do desempenho em cada categoria de pneumonia, diferenciando-a de outras estratégias que avaliam o desempenho global. Isso garante uma maior confiabilidade no suporte ao diagnóstico médico, além de destacar a eficácia do modelo em contextos clínicos reais, onde a classificação de pneumonia pode ser mais complexa do que uma simples classificação binária.

A discussão apresentada nos resultados confirmam a eficácia do modelo proposto para a classificação de pneumonia com base em imagens de raios-x, considerando diferentes cenários de comparação, como pneumonia bacteriana, viral e ausência de pneumonia. A abordagem de análise se complementa como forma de auxílio ao médico, ajudando a separar os casos de maiores tendências a terem a patologia e agilizando os possíveis tratamentos, é crucial destacar a importância da separação dos casos para o diagnóstico e tratamento adequado das pneumonias. Pneumonias bacterianas e virais requerem abordagens terapêuticas distintas: enquanto as primeiras geralmente respondem bem a antibióticos, as virais podem demandar tratamentos antivirais específicos ou apenas suporte clínico. A correta identificação da etiologia não só otimiza o manejo clínico como também reduz o uso inadequado de medicamentos, contribuindo para a prevenção da resistência antimicrobiana. Nesse sentido, o modelo proposto pode atuar como uma ferramenta auxiliar relevante, ajudando médicos a diferenciar rapidamente os tipos de pneumonia e promovendo decisões mais informadas.

Embora os resultados mostrem espaço para melhorias, como em sensibilidade e precisão, o uso de técnicas adicionais de aprendizado de máquina e aumento de dados pode permitir uma maior precisão. Em estudos futuros podemos explorar a aplicação de redes neurais mais complexas, a diversificação do *Dataset* com diferentes tipos de imagens para melhorar o desempenho do diagnóstico automático de patologias, e melhorar a visualização para o usuário.

REFERÊNCIAS

- AGGARWAL, C. C. **Neural Networks and Deep Learning: A Textbook**. [S.l.]: Springer, 2018.
- AQUINO, C. P.; FONSECA, A. U. d.; OLIVEIRA, L. L. G. d. Avaliação de classificadores no diagnóstico de câncer de mama. **J. health inform**, [S.l.], p.915–926, 2016.
- ARAÚJO, J. S.; HORNUNG, J. A. Inteligência artificial no diagnóstico de doenças: a responsabilidade civil do médico em caso de erro de diagnóstico. **THEMIS: Revista da Esmec**, [S.l.], v.20, n.1, p.113–145, 2022.
- ARRUDA, M. I. M. et al. **O laudo médico-legal como fonte de informação e seu papel social**. [S.l.]: Instituto Brasileiro de Informação em Ciência e Tecnologia/Universidade . . . , 2000.
- ASSUNÇÃO, R. G.; PEREIRA, W. A.; ABREU, A. G. Pneumonia bacteriana: aspectos epidemiológicos, fisiopatologia e avanços no diagnóstico. **Rev Inv Biomédica**, [S.l.], v.10, n.1, p.83–91, 2018.
- AYAN, E.; ÜNVER, H. M. Diagnosis of pneumonia from chest X-ray images using deep learning. In: SCIENTIFIC MEETING ON ELECTRICAL-ELECTRONICS & BIOMEDICAL ENGINEERING AND COMPUTER SCIENCE (EBBT), 2019., 2019. **Anais. . .** [S.l.: s.n.], 2019. p.1–5.
- BENGIO, Y. Learning Deep Architectures for AI. **Foundations and Trends® in Machine Learning**, [S.l.], v.2, n.1, p.1–127, 2009.
- BHANDARY, A.; PRABHU, G. A.; RAJINIKANTH, V.; THANARAJ, K. P.; SATAPATHY, S. C.; ROBBINS, D. E.; SHASKY, C.; ZHANG, Y.-D.; TAVARES, J. M. R.; RAJA, N. S. M. Deep-learning framework to detect lung abnormality—A study with chest X-Ray and lung CT scan images. **Pattern Recognition Letters**, [S.l.], v.129, p.271–278, 2020.
- BISHOP, C. M. **Neural networks for pattern recognition**. [S.l.]: Oxford university press, 1995.
- BRANCO, S. **O domínio público no direito autoral brasileiro: uma obra em domínio público**. [S.l.]: Lumen Juris, 2011.
- BREIMAN, L. Random forests. **Machine learning**, [S.l.], v.45, p.5–32, 2001.
- CURRY, T. S.; DOWDEY, J. E.; MURRY, R. C. **Christensen’s physics of diagnostic radiology**. [S.l.]: Lippincott Williams & Wilkins, 1990.

- DÍAZ NOVÁS, J.; GALLEGU MACHADO, B.; LEÓN GONZÁLEZ, A. El diagnóstico médico: bases y procedimientos. **Revista Cubana de Medicina General Integral**, [S.l.], v.22, n.1, 2006.
- FELSON, B.; GOODMAN, L. R. **Felson's Principles of Chest Roentgenology**. 4th.ed. [S.l.]: Elsevier, 2014.
- FERRAZ, R. d. O.; OLIVEIRA-FRIESTINO, J. K.; FRANCISCO, P. M. S. B. Pneumonia mortality trends in all Brazilian geographical regions between 1996 and 2012. **Jornal Brasileiro de Pneumologia**, [S.l.], v.43, n.04, p.274–279, 2017.
- FREUND, Y.; SCHAPIRE, R. E. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. **Journal of Computer and System Sciences**, [S.l.], v.55, n.1, p.119–139, 1997.
- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). **The annals of statistics**, [S.l.], v.28, n.2, p.337–407, 2000.
- GERON, A. **Mãos à obra: aprendizado de maquina com Scikit-Learn, Keras & Tensorflow**. [S.l.]: Rio de Janeiro: Alta Books, 2021.
- GOLDMAN, L.; AUSIELLO, D. **Cecil Medicina**. [S.l.]: Elsevier, 2016.
- GONZALEZ, R. C.; WOODS, R. E. Processamento digital de imagem. **Pearson, ISBN-10: 8576054019**, [S.l.], v.10, p.11–27, 2010.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep Learning**. [S.l.]: MIT Press, 2016.
- GULLI, A.; PAL, S. **Deep learning with Keras**. [S.l.]: Packt Publishing Ltd, 2017.
- HANIFI, S. M.; MAHMOOD, S. S.; BHUIYA, A. Cause-specific mortality and socioeconomic status in Chakaria, Bangladesh. **Global health action**, [S.l.], v.7, n.1, p.25473, 2014.
- HINTON, G. E.; OSINDERO, S.; TEH, Y.-W. A fast learning algorithm for deep belief nets. **Neural computation**, [S.l.], v.18, n.7, p.1527–1554, 2006.
- HOFMANN, T.; SCHÖLKOPF, B.; SMOLA, A. J. Kernel methods in machine learning. **Annals of Statistics**, [S.l.], v.36, n.3, p.1171–1220, 2008.
- IBRAHIM, A. U.; OZSOZ, M.; SERTE, S.; AL-TURJMAN, F.; YAKOI, P. S. Pneumonia classification using deep learning from chest X-ray images during COVID-19. **Cognitive computation**, [S.l.], p.1–13, 2021.
- JAIN, R.; NAGRATH, P.; KATARIA, G.; KAUSHIK, V. S.; HEMANTH, D. J. Pneumonia detection in chest X-ray images using convolutional neural networks and transfer learning. **Measurement**, [S.l.], v.165, p.108046, 2020.

- JAMESON, J. L.; FAUCI, A. S.; KASPER, D. L.; HAUSER, S. L.; LONGO, D. L.; LOS-CALZO, J. **Medicina Interna de Harrison-2 Volumes-20**. [S.l.]: McGraw Hill Brasil, 2020.
- JAUHAR, M. H. O.; THOMES, C. R. Inteligência Artificial em Imagens Médicas: Impactos e Desafios. **Cognitus Interdisciplinary Journal**, [S.l.], v.2, n.1, p.160–171, February 2025.
- KAUCZOR, H.-U.; KANNE, J. P.; LYNCH, D. A. The Fleischner Society: 2017 Update on the Glossary of Terms for Thoracic Imaging. **Radiology**, [S.l.], v.284, n.2, p.545–560, 2017.
- KERMANY, D. S.; GOLDBAUM, M.; CAI, W.; VALENTIM, C. C.; LIANG, H.; BAXTER, S. L.; MCKEOWN, A.; YANG, G.; WU, X.; YAN, F. et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. **cell**, [S.l.], v.172, n.5, p.1122–1131, 2018.
- KERMANY, D.; ZHANG, K.; GOLDBAUM, M. **Large Dataset of Labeled Optical Coherence Tomography (OCT) and Chest X-ray Images**.
- KUMAR, V.; ABBAS, A. K.; ASTER, J. C. **Pathologic Basis of Disease**. [S.l.]: Elsevier, 2020.
- LAFETÁ, B. O.; SILVA, F. F. da; SANTOS, M. A. dos; PIMENTA, I. A.; FONTAN, I. d. C. I.; FONSECA, N. R.; SARTORI, C. J. Modelagem morfométrica de *Licania tomentosa* (Benth.) por regressão logística e máquinas vetor de suporte. **Scientia Plena**, [S.l.], v.16, n.6, 2020.
- LEE, K.-Y. Pediatric respiratory infections by *Mycoplasma pneumoniae*. **Expert review of anti-infective therapy**, [S.l.], v.6, n.4, p.509–521, 2008.
- LEE, K.-Y.; YOUN, Y.-S.; LEE, J.-W.; KANG, J.-H. **Pneumonia por micoplasma, pneumonia bacteriana e pneumonia viral**. [S.l.]: SciELO Brasil, 2010. 448–450p. v.86.
- MATOSO, L. M. L.; CASTRO, C. H. A. de. Indissociabilidade clínica e epidemiológica da pneumonia. **CATUSSABA-ISSN 2237-3608**, [S.l.], v.2, n.2, p.11–24, 2013.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. **The bulletin of mathematical biophysics**, [S.l.], v.5, p.115–133, 1943.
- MINSKY, M.; PAPERT, S. An introduction to computational geometry. **Cambridge tiass., HIT**, [S.l.], v.479, n.480, p.104, 1969.
- MURRAY, J. F.; NADEL, J. A. **Murray & Nadel's Textbook of Respiratory Medicine**. 7th.ed. [S.l.]: Elsevier, 2021.
- PANG, B.; NIJKAMP, E.; WU, Y. N. Deep learning with tensorflow: A review. **Journal of Educational and Behavioral Statistics**, [S.l.], v.45, n.2, p.227–248, 2020.

- POLIKAR, R. Ensemble based systems in decision making. **IEEE Circuits and systems magazine**, [S.l.], v.6, n.3, p.21–45, 2006.
- RAHAT, B. **Principles of Laboratory Diagnosis**. [S.l.]: Springer, 2021.
- RAKEL, D. **Textbook of Family Medicine**. [S.l.]: Elsevier, 2022.
- REZAAEE, K.; BADIEI, A.; MESHGINI, S. A hybrid deep transfer learning based approach for COVID-19 classification in chest X-ray images. In: IRANIAN CONFERENCE ON BIOMEDICAL ENGINEERING (ICBME), 2020., 2020. **Anais. . .** [S.l.: s.n.], 2020. p.234–241.
- ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain [J]. **Psychological review**, [S.l.], v.65, n.6, p.386–408, 1958.
- SHANMUGAMANI, R. **Deep Learning for Computer Vision**. [S.l.]: Packt Publishing, 2018.
- SILVA, A. M. Marques da; PATROCÍNIO, A. C.; SCHIABEL, H. Processamento e Análise de Imagens Médicas. **Revista Brasileira de Física Médica**, [S.l.], v.13, n.1, p.34–48, September 2019.
- SIMON, G. **Radiology of Chest Diseases**. 3rd.ed. [S.l.]: W. B. Saunders Company, 1997.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. **arXiv preprint arXiv:1409.1556**, [S.l.], 2014.
- TISIOLOGIA, S. B. de Pneumologia e Diretrizes para o Diagnóstico e Tratamento das Pneumonias Adquiridas na Comunidade. **Jornal Brasileiro de Pneumologia**, [S.l.], v.47, n.2, p.e20210022, 2021.
- TSIKNAKIS, N.; TRIVIZAKIS, E.; VASSALOU, E. E.; PAPADAKIS, G. Z.; SPANDIDOS, D. A.; TSATSAKIS, A.; SÁNCHEZ-GARCÍA, J.; LÓPEZ-GONZÁLEZ, R.; PAPANIKOLAOU, N.; KARANTANAS, A. H. et al. Interpretable artificial intelligence framework for COVID-19 screening on chest X-rays. **Experimental and therapeutic medicine**, [S.l.], v.20, n.2, p.727–735, 2020.
- VAN ROSSUM, G. et al. Python Programming Language. In: USENIX ANNUAL TECHNICAL CONFERENCE, 2007. **Anais. . .** [S.l.: s.n.], 2007. v.41, n.1, p.1–36.
- VAPNIK, V. **The nature of statistical learning theory**. [S.l.]: Springer science & business media, 1999.
- ZHONG, G.; WANG, L.-N.; LING, X.; DONG, J. An overview on data representation learning: From traditional feature learning to recent deep learning. **The Journal of Finance and Data Science**, [S.l.], v.2, n.4, p.265–278, 2016.