



Universidade de Brasília – UnB  
Campus Gama – FGA  
Programa de Pós-Graduação em Engenharia Biomédica

**IDENTIFICAÇÃO DE LESÕES DA LARINGE  
A PARTIR DE IMAGEAMENTO POR ENDOSCOPIA DE BANDA ESTREITA  
UTILIZANDO REDES NEURAIS ARTIFICIAIS E PROGRAMAÇÃO VISUAL**

**ANA LUIZA MORAES FERNANDES DA COSTA**

Orientador: PROFESSOR DR. GERARDO ANTONIO IDROBO PIZO



UNB – UNIVERSIDADE DE BRASÍLIA

FGA – FACULDADE GAMA



**IDENTIFICAÇÃO DE LESÕES DA LARINGE  
A PARTIR DE IMAGEAMENTO POR ENDOSCOPIA DE BANDA ESTREITA  
UTILIZANDO REDES NEURAS ARTIFICIAIS E PROGRAMAÇÃO VISUAL**

**ANA LUIZA MORAES FERNANDES DA COSTA**

ORIENTADOR: PROFESSOR DR. GERARDO ANTONIO IDROBO PIZO

DISSERTAÇÃO DE MESTRADO EM  
ENGENHARIA BIOMÉDICA

PUBLICAÇÃO: 0163A/2023

BRASÍLIA/DF, MAIO DE 2023

**UNB – UNIVERSIDADE DE BRASÍLIA**  
**FGA – FACULDADE GAMA**  
**PROGRAMA DE PÓS-GRADUAÇÃO**

**IDENTIFICAÇÃO DE LESÕES DA LARINGE**  
**A PARTIR DE IMAGEAMENTO POR ENDOSCOPIA DE BANDA ESTREITA**  
**UTILIZANDO REDES NEURAI ARTIFICIAIS E PROGRAMAÇÃO VISUAL**

ANA LUIZA MORAES FERNANDES DA COSTA

DISSERTAÇÃO DE MESTRADO SUBMETIDA AO PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA BIOMÉDICA DA UNIVERSIDADE DE BRASÍLIA, COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM ENGENHARIA BIOMÉDICA

APROVADA POR:

---

Professor Dr. Gerardo Antonio Idrobo Pizo  
(Orientador)

---

Professora Dra. Luciana Correia Alves  
(Examinador externo) UNICAMP

---

Professor Dr. Cristiano Jacques Miosso  
(Examinador interno) UnB

---

Professor Dr. Renato Coral Sampaio  
(Examinador suplente) UnB

**FICHA CATALOGRÁFICA**

COSTA, ANA LUIZA MORAES FERNANDES DA

Identificação de Lesões da Laringe

a Partir de Imageamento por Endoscopia de Banda Estreita

Utilizando Redes Neurais Artificiais e Programação Visual

[Distrito Federal], 2023.

69p., 210 × 297 mm (FGA/UnB Gama, Mestrado em Engenharia Biomédica, 2023).

Dissertação de Mestrado em Engenharia Biomédica, Faculdade UnB Gama, Programa de Pós-Graduação em Engenharia Biomédica.

- |                            |                      |
|----------------------------|----------------------|
| 1. Inteligência Artificial | 2. Câncer de Laringe |
| 3. Capsule Network         | 4. NBI               |
| I. FGA UnB/UnB.            | II. Título (série)   |

**REFERÊNCIA**

COSTA, ANA LUIZA MORAES FERNANDES DA (2023). Identificação de Lesões da Laringe a Partir de Imageamento por Endoscopia de Banda Estreita Utilizando Redes Neurais Artificiais e Programação Visual. Dissertação de mestrado em engenharia biomédica, Publicação 0163A/2023, Programa de Pós-Graduação, Faculdade UnB Gama, Universidade de Brasília, Brasília, DF, 69p.

**CESSÃO DE DIREITOS**

AUTOR: Ana Luiza Moraes Fernandes da Costa

TÍTULO: Identificação de Lesões da Laringe

a Partir de Imageamento por Endoscopia de Banda Estreita

Utilizando Redes Neurais Artificiais e Programação Visual

GRAU: Mestre

ANO: 2023

É concedida à Universidade de Brasília permissão para reproduzir cópias desta dissertação de mestrado e para emprestar ou vender tais cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte desta dissertação de mestrado pode ser reproduzida sem a autorização por escrito do autor.

---

[analuzamoraesfc@gmail.com](mailto:analuzamoraesfc@gmail.com)

Brasília, DF – Brasil

*"Pois Dele, por Ele e para Ele são todas as coisas. A Ele seja a glória para sempre!".*  
(Romanos 11:36)

*Dedico este trabalho à Deus, à minha amada mãe, Gilda, ao meu esposo, Igor, e à  
minha querida filha, Emily.*

## **AGRADECIMENTOS**

Agradeço primeiramente à Deus, por ter me sustentado e me dado forças durante essa jornada. Agradeço pelo dom da vida, por tudo que sou e que tenho, tudo devo a Ele. Portanto, agradeço a Deus, que abriu os caminhos e me deu a oportunidade de chegar até esse momento tão importante em minha vida. A Ele, toda Honra e toda Glória.

Agradeço também a minha amada mãe, Gilda, por ser minha maior incentivadora, a pessoa que sempre me apoiou e acreditou em mim, e é a grande responsável pela realização desse sonho. É para você que dedico esse trabalho.

Agradeço ao meu esposo, Igor, pela paciência e pela parceria durante o curso e também à minha amada filha Emily, que me inspira e me motiva a me tornar uma pessoa melhor a cada dia.

Ao professor Dr. Gerardo Pizo, agradeço pela atenção, pelas orientações, pelo apoio e pela disponibilidade em compartilhar o conhecimento, que tornaram possível essa entrega.

Agradeço também aos meus familiares, meus sogros e meus líderes Giselle e Marcus, que oraram por mim e me deram a força espiritual necessária para alcançar essa conquista.

Ao Programa de Pós-Graduação em Engenharia Biomédica da Universidade de Brasília e seu corpo docente, pelo ensino e pesquisa de excelência oportunizado durante todo o curso.

## RESUMO

A presença de determinados tipos de lesões na mucosa da laringe pode sinalizar o desenvolvimento de carcinoma espinocelular laríngeo em estágio inicial, que surge a partir do desenvolvimento anormal das células escamosas que compõem a camada interna do revestimento da laringe. Essa doença corresponde a 98% das ocorrências de tumores malignos que acometem a laringe e 3% de todos os casos de câncer, frequentemente levando à óbito quando diagnosticada em estágios avançados.

A detecção precoce de lesões sugestivas de desenvolvimento de tecido pré-cancerígeno ou de câncer da laringe em estágio inicial pode ser uma tarefa desafiadora mesmo para médicos experientes. A literatura científica existente aponta que modelos baseados em aprendizado de máquina têm sido utilizados para auxiliar o diagnóstico de câncer de laringe, alcançando resultados relevantes, sendo consideradas no estado da arte para a detecção de lesões da laringe as arquiteturas de redes neurais artificiais convolucionais profundas, com precisão reportada de até 98%. Entretanto, a implementação de modelos de *machine learning* frequentemente exige conhecimentos avançados em linguagem de programação, tornando complexa a tarefa de treinamento e desenvolvimento dessas soluções. Há também modelos de redes neurais artificiais mais recentes, ainda não explorados para a classificação de imagens contendo lesões da laringe.

O presente estudo teve por objetivo desenvolver novos métodos baseados em inteligência artificial para classificar lesões da laringe de forma automatizada, a partir de imagens digitais obtidas pela técnica de imagens de banda estreita.

Relativamente à metodologia adotada, foram utilizadas 1.320 imagens digitais de tecido da laringe, saudável e em estágio inicial de câncer, obtidas por meio de laringoscopia, subdividas em 4 classes: tecido saudável, tecido com leucoplasia, tecido com vasos hipertróficos e tecido com alças capilares intrapapilares. Desse conjunto, foram segregadas 132(10%) para compor o conjunto de testes. Foram implementados cinco modelos de aprendizagem de máquina diferentes, baseados em redes neurais convolucionais profundas (CNN) utilizando programação visual e redes de cápsulas (CapsNet), para classificar as 132 imagens: VGG16, VGG19, Inception V3, CapsNet sem *data augmentation* e CapsNet combinado a uma técnica de *data augmentation* utilizando imagens geradas sinteticamente por uma rede generativa adversária (GAN).

As arquiteturas de rede neural foram implementadas por meio do software livre Orange Data Mining e da plataforma computacional Google Colab, em linguagem Python, com a utilização das bibliotecas Keras, OpenCV e Tensorflow.

O uso de GAN para ampliar o conjunto de treinamento com imagens geradas sinteticamente melhorou o desempenho do classificador de cápsula na classificação de todas os tipos de lesões. O classificador CapsNet com aumento de dados por GAN alcançou sensibilidade,

precisão e F1-Score médios de 94,7%, correspondente ao segundo melhor desempenho entre os modelos estudados, contudo ainda apresentou indícios de sobreajuste. O melhor desempenho geral foi obtido com o classificador CNN Inception V3, que obteve 97% de sensibilidade, precisão e F1-Score, utilizando software de programação visual.

Este estudo contribuiu para o avanço do conhecimento em tecnologias relacionadas à saúde ao explorar novos métodos para o auxílio do diagnóstico de câncer da laringe. A solução de inteligência artificial baseada em Rede Neural Convolutiva (do inglês *Convolutional Neural Network*) (CNN) com programação visual proposta alcançou resultados competitivos, permitindo detectar as lesões da laringe em todas as imagens apresentadas ao modelo, com a vantagem de ser implementada em ferramenta de fácil utilização, com menor custo de recursos computacionais e por não exigir conhecimentos avançados em programação de modelos de *machine learning*, comparativamente aos demais métodos já explorados.

**Palavras-chave:** inteligência artificial, capsule networks, carcinoma espinocelular da laringe, imagens de banda estreita.

## ABSTRACT

The presence of certain types of lesions in the laryngeal mucosa can signal the development of early-stage laryngeal squamous cell carcinoma, which arises from the abnormal development of the squamous cells that make up the inner layer of the laryngeal lining. This disease corresponds to 98% of the occurrences of malignant tumors that affect the larynx and 3% of all cases of cancer, often leading to death when diagnosed in advanced stages.

Early detection of lesions suggestive of the development of precancerous tissue or early stage laryngeal cancer can be a challenging task even for experienced physicians. The existing scientific literature points out that models based on machine learning have been used to aid the diagnosis of laryngeal cancer, achieving relevant results, considering the state-of-the-art architecture of deep convolutional artificial neural networks in the detection of laryngeal lesions, with reported accuracy of up to 98%. However, the implementation of *machine learning* models often requires advanced knowledge in programming language, making the task of training and developing these solutions complex. There are also more recent models of artificial neural networks, not yet explored for the classification of images containing laryngeal lesions.

The present study aimed to develop methods based on artificial intelligence to classify laryngeal lesions in an automated way, from digital images obtained by the narrowband imaging technique.

Regarding the adopted methodology, 1,320 digital images of laryngeal tissue, healthy and in an early stage of cancer, obtained through laryngoscopy, were used, subdivided into 4 classes: healthy tissue, tissue with leukoplakia, tissue with hypertrophic vessels and tissue with capillary loops intrapapillary. From this set, 132(10%) were segregated to compose the test set. Five different machine learning models, based on deep convolutional neural networks (CNN) and capsule networks (CapsNet), were implemented to classify the 132 images: VGG16, VGG19, Inception V3, CapsNet without *data augmentation* and CapsNet combined with an *data augmentation* technique using images generated synthetically by an adversarial generative network (GAN).

The neural network architectures were implemented using the free software ORANGE Data mining and the Google Colab computational platform, in Python language, using the Keras, OpenCV and Tensorflow libraries.

Using GAN to augment the training set with synthetically generated images improved the performance of the capsule classifier in classifying all types of injuries. The CapsNet classifier with data increase by GAN reached average recall, precision and F1-Score of

94.7%, reaching the second best performance among the studied models. The best performance was obtained with the CNN Inception V3 classifier, which obtained 97% recall, precision and F1-Score, using visual programming software.

This study contributed to the advancement of knowledge in health-related technologies by exploring new methods to aid in the diagnosis of laryngeal cancer. The artificial intelligence solution based on Rede Neural Convolutacional (do inglês *Convolutional Neural Network*) (CNN) with the proposed visual programming reached competitive results, allowing the detection of laryngeal lesions in all images presented to the model, with the advantage of being implemented in an easy-to-use tool, with a lower cost of computational resources and because it does not require advanced knowledge in programming *machine learning* models, compared to the other methods already explored.

**Keywords:** artificial intelligence, capsule networks, laryngeal squamous cell carcinoma, narrowband imaging.

## SUMÁRIO

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Contextualização e Definição do Problema . . . . .	1
1.2	Objetivos . . . . .	3
1.2.1	Objetivo Geral . . . . .	3
1.2.2	Objetivos Específicos . . . . .	4
1.3	Justificativa . . . . .	4
1.4	Estrutura da Dissertação . . . . .	5
<b>2</b>	<b>Fundamentação Teórica</b>	<b>7</b>
2.1	Carcinoma Espinocelular e Lesões da Laringe . . . . .	7
2.2	Processamento Digital de Imagens . . . . .	12
2.2.1	Etapas do Processamento Digital de Imagens . . . . .	13
2.3	Mineração de Dados e Classificação de Imagens . . . . .	16
2.3.1	Mineração de Dados . . . . .	16
2.3.2	Redes Neurais Artificiais e Classificação de Imagens Médicas . . . . .	17
2.3.3	<i>Data Augmentation</i> . . . . .	20
2.3.4	Ferramentas de Mineração de Dados e Análise de Imagens . . . . .	22
2.4	Trabalhos Relacionados . . . . .	24
2.5	Considerações Finais . . . . .	28
<b>3</b>	<b>Materiais e Métodos</b>	<b>29</b>
3.1	Aquisição do Banco de Imagens . . . . .	30
3.2	<i>Data Aumentation</i> com Rede Generativa DCGAN . . . . .	31
3.3	Mineração dos Dados e Classificação das Imagens . . . . .	35

3.4	Métricas para Avaliação dos Resultados . . . . .	36
<b>4</b>	<b>Resultados e Discussão</b>	<b>39</b>
4.1	Resultado da Geração das Imagens . . . . .	39
4.2	Resultados obtidos pelo Classificador CNN arquitetura VGG-16 . . . . .	40
4.3	Resultados obtidos pelo Classificador CNN arquitetura VGG-19 . . . . .	41
4.4	Resultados obtidos pelo Classificador CNN arquitetura Inception V3 . . . . .	42
4.5	Resultados obtidos com CapsNet . . . . .	43
4.6	Considerações Finais . . . . .	48
<b>5</b>	<b>Conclusão</b>	<b>50</b>
<b>6</b>	<b>Limitações e Trabalhos Futuros</b>	<b>52</b>
	<b>Lista de Referências</b>	<b>53</b>
	<b>Apêndices</b>	<b>60</b>
	Apêndice A - Códigos Implementados . . . . .	60
	Apêndice B - <i>Slides</i> da Apresentação . . . . .	61

## LISTA DE TABELAS

2.1	Quantidade de Artigos Relacionados Publicados . . . . .	25
3.1	Organização do Dataset sem Aumento de Dados . . . . .	31
3.2	Arquitetura da rede geradora DCGAN . . . . .	32
3.3	Arquitetura da rede discriminativa DCGAN . . . . .	33
4.1	Resultado da Ampliação do Banco de Imagens com DCGAN . . . . .	40
4.2	Sensibilidade na classificação com CapsNet, por classe . . . . .	47
4.3	Precisão na classificação com CapsNet, por classe . . . . .	47
4.4	F1-Score na classificação com CapsNet, por classe . . . . .	48
4.5	Resumo dos Resultados . . . . .	48

## LISTA DE QUADROS

2.1	Trabalhos Relacionados - Base <i>Scopus</i> . . . . .	26
2.2	Trabalhos Relacionados - Base <i>Science Direct</i> . . . . .	27
2.3	Trabalhos Relacionados - Base <i>Web of Science</i> . . . . .	27
4.1	Análise dos resultados junto aos trabalhos relacionados . . . . .	49

## LISTA DE FIGURAS

2.1	Ilustração da diferença entre mucosa saudável, com displasia e com carcinoma invasivo. . . . .	8
2.2	Leucoplasia na laringe. . . . .	8
2.3	Padrão de Alças Capilares Intrapapilares. . . . .	9
2.4	Exemplo de capturas em NBI de tecido saudável da laringe. . . . .	9
2.5	Classificação dos Tipos de Lesões em Alças Capilares Intrapapilares na Laringe. . . . .	11
2.6	Imagem digital monocromática representada nas coordenadas X e Y. . .	12
2.7	Representação do Espaço de Cores RGB. . . . .	13
2.8	Etapas do Processamento Digital de Imagens. . . . .	14
2.9	Exemplo de segmentação de imagem. . . . .	15
2.10	Estrutura de um neurônio artificial. . . . .	17
2.11	Estrutura típica de uma rede neural artificial. . . . .	18
2.12	Estrutura de uma rede CNN em dois estágios . . . . .	19
2.13	Esquema de arquitetura de rede em cápsula. . . . .	20
2.14	Data Augmentation com Transformações Geométricas . . . . .	21
2.15	Estrutura de uma rede GAN. . . . .	22
2.16	Modelo de análise de imagens implementado no software Orange. . . . .	23
3.1	Fluxograma com etapas adotadas na metodologia . . . . .	29
3.2	Amostra do dataset de imagens do tecido da laringe. . . . .	30
3.3	Curva perda por época obtida pela rede DCGAN. . . . .	34
3.4	Implementação dos Classificadores CNN no Software Orange. . . . .	36
3.5	Exemplo de implementação de um modelo CapsNet utilizando a biblioteca Keras em python . . . . .	37

3.6	Exemplo de Matriz de Confusão. . . . .	38
4.1	Amostra de resultados da DCGAN. . . . .	39
4.2	Matriz de Confusão do Classificador CNN VGG 16. . . . .	40
4.3	Matriz de Confusão do Classificador CNN VGG 19. . . . .	41
4.4	Matriz de Confusão do Classificador CNN Inception V3. . . . .	42
4.5	Matriz de Confusão do Classificador CapsNet. . . . .	43
4.6	Evolução da curva de perda no Treinamento e Validação por época para a classificação com CapsNet sem aumento de dados. . . . .	44
4.7	Acurácia obtida no Treinamento e Validação por época na classificação com CapsNet. . . . .	44
4.8	Evolução da curva de perda no Treinamento e Validação por época para a classificação com CapsNet e aumento de dados . . . . .	45
4.9	Acurácia obtida no Treinamento e Validação por época na classificação com CapsNet e aumento de dados. . . . .	45
4.10	Matriz de Confusão do Classificador CapsNet com Aumento de Dados. . . . .	46

## LISTA DE NOMENCLATURAS E ABREVIACÕES

**API** Interface de Programação de Aplicações, do inglês *Application Programming Interface*

**CapsNet** Rede Neural em Cápsula

**CECL** Carcinoma Espinocelular da Laringe

**CNN** Rede Neural Convolutacional (do inglês *Convolutional Neural Network*)

**DCGAN** Rede Generativa Adversária Convolutacional Profunda (do inglês *Deep Convolutional Generative Adversarial Network*)

**EC** Endoscopia de Contato

**GAN** Rede Generativa Adversária (do inglês *Generative Adversarial Network*)

**GPU** Unidade Gráfica de Processamento (do inglês *Graphical Processing Unit*)

**IA** Inteligência Artificial

**IQR** Intervalo Interquartil

**kNN** k-Nearest Neighbours

**NBI** Imagem de Banda Estreita (do inglês *Narrow Band Imaging*)

**PDI** Processamento Digital de Imagens

**RF** Floresta Aleatória (do inglês *Random Forest*)

**RNA** Rede Neural Artificial

**SVM** Máquina de Vetores de Suporte (do inglês *Support Vector Machine*)

**ROC** Curva Característica de Operação do Receptor

# 1 INTRODUÇÃO

## 1.1 CONTEXTUALIZAÇÃO E DEFINIÇÃO DO PROBLEMA

A laringe é um órgão musculocartilaginoso, multifuncional, localizado na região infra-hióidea. As principais funções da laringe consistem em conduzir o ar durante a respiração e proteger as vias aéreas inferiores. Esse órgão também possui função esfínteriana, evitando a entrada de água e alimentos; e inibindo a saída de ar dos pulmões durante esforços fisiológicos (Nemetz *et al.*, 2005).

Além disso, a laringe tem papel fundamental na produção da voz humana. A produção da voz em seres humanos depende de um conjunto de diversos mecanismos, tais como a ressonância do trato vocal, as pressões subglotal e supraglotal, as características biomecânicas dos tecidos que revestem a laringe e a oscilação das pregas vocais durante a fonação (Scalassara *et al.*, 2009).

O câncer de laringe corresponde a cerca de 3% de todos os casos de neoplasia maligna no mundo (Dogantemur *et al.*, 2020). Estima-se um total de mais 184 mil ocorrências dessa neoplasia no ano de 2020 (WHO, 2020). Em 2020, o número de casos novos de câncer de laringe estimados para o Brasil foi de 6.470 em homens e de 1.180 em mulheres. Nesse ano, foram registrados 4.450 óbitos decorrentes da doença no país (INCA, 2020). Para 2023, a taxa de incidência de câncer de laringe estima um risco de 6,21 casos novos a cada 100 mil homens e de 1,09 casos novos a cada 100 mil mulheres no Brasil (INCA, 2023).

As manifestações clínicas do câncer laríngeo mais comuns se referem à apresentação de dificuldade na deglutição ou rouquidão na voz persistentes, acompanhados de nódulos cervicais ou não (Hermans, 2006). Em que pese os sintomas precoces, o carcinoma epinocelular da laringe ainda é frequentemente diagnosticado em estágio avançado (Amar *et al.*, 2010).

O tratamento dessa neoplasia é por vezes invasivo, incluindo laringectomias totais somadas à quimioterapia e radioterapia, o que pode contribuir para exacerbar os sintomas da doença (Maciel *et al.*, 2010), impactando significativamente a qualidade de vida dos indivíduos. Assim, a identificação do câncer de laringe em estágio inicial revela-se

fundamental para um melhor prognóstico da doença, visto que o diagnóstico precoce é um importante fator para a redução da taxa de mortalidade e para a preservação da anatomia da laringe e das pregas vocais (Unger *et al.*, 2015).

A técnica endoscópica de Imagem de Banda Estreita (do inglês *Narrow Band Imaging*) (NBI) constitui uma moderna ferramenta para o diagnóstico do câncer de laringe, e consiste em uma abordagem que utiliza luz para aumentar os padrões neoangiogênicos associados com lesões pré-neoplásicas e cancerosas da mucosa parte superior do trato aerodigestivo (Piazza *et al.*, 2012).

Os principais pontos de atenção sugestivos de neoplasia maligna da laringe, passíveis de identificação pela avaliação de imagens obtidas por NBI, residem na presença de vasos hipertróficos com alças vesiculares, e na existência de leucoplasia nas pregas vocais.

Apesar dos avanços nos procedimentos diagnósticos, a identificação de tecidos com lesões suspeitas durante o exame de endoscopia persiste um desafio para os especialistas clínicos, visto que, por vezes, alterações mais sutis na mucosa podem ser imperceptíveis ao olho humano.

A utilização de imagens digitais tem se tornado cada vez mais frequente em áreas como diagnósticos médicos, vigilância e monitoramento por câmeras de segurança e satélites, entre outros (Liu *et al.*, 2017).

Ferramentas de auxílio ao diagnóstico de patologias baseadas em inteligência artificial têm sido amplamente aplicadas à medicina (Lobo, 2017). Assim, métodos de aprendizado de máquina têm sido utilizados para classificação de diferentes modalidades de imagens médicas (Santos *et al.*, 2020), agregando maior assertividade na identificação de doenças.

Redes neurais convolucionais consistem em uma tecnologia amplamente utilizada com o objetivo de realizar tarefas de reconhecimento de imagens. A literatura científica reporta que modelos baseados em aprendizado de máquina tem atingindo resultados significativos a detecção de lesões da laringe, com precisão de até 98% Moccia2017, sendo consideradas como estado da arte as arquitetura de redes neurais artificiais convolucionais profundas.

Contudo, o desenvolvimento de soluções de *machine learning* é uma atividade que requer conhecimentos em programação, consistindo em tecnologia pouco acessível para usuários que não detém formação na área de computação. Não obstante os relevantes resultados obtidos por Rede Neural Convolucional (do inglês *Convolutional Neural Network*) (CNN), existem abordagens mais recentes de redes neurais artificiais, como as redes neurais de cápsula, ainda não exploradas no contexto de classificação de imagens contendo lesões da laringe.

As redes de cápsula consistem em uma arquitetura mais recente que objetiva superar algumas limitações das redes convolucionais convencionais, a medida em que se diferen-

ciam destas pela capacidade de preservar as informações de posição e a relação hierárquica entre os objetos e a composição completa da imagem.

O presente estudo se dedicou a responder as seguintes questões de pesquisa:

- Qual o desempenho de um modelo de aprendizagem de máquina supervisionado baseado em arquitetura de rede de cápsula na tarefa de classificar lesões da laringe em imagens obtidas por NBI, comparativamente às principais arquiteturas de Rede Neural Artificial (RNA) exploradas na literatura científica?
- A adição de imagens médicas sintéticas geradas por Rede Generativa Adversária (do inglês *Generative Adversarial Network*) (GAN) à base de treinamento influencia positivamente o desempenho do classificador baseado em rede de cápsula?
- Qual o desempenho de uma solução baseada em inteligência artificial, implementada com programação visual, na tarefa de classificação de lesões em imagens da mucosa laríngea?

Esta pesquisa propôs explorar uma arquitetura de classificador baseada em redes neurais artificiais em cápsula para automatizar a classificação de lesões indicativas de carcinoma espinocelular da laringe, a partir da avaliação de imagens do tecido da mucosa laríngea obtidos pela técnica de NBI. A metodologia utilizada nesta pesquisa contemplou adoção de recurso de *data augmentation* através da geração de imagens médicas sintéticas por redes generativas adversárias, para ampliar a quantidade e a variabilidade das bases de dados, com a expectativa de melhoramento da performance do classificador.

Este estudo também se dedicou à implementação de classificadores fundamentados em redes do tipo CNN, especificamente constituídos sob as arquiteturas Inception, VGG16 e VGG19, por meio de uma ferramenta de programação visual de acesso livre, que possui interface gráfica amigável para usuários leigos em computação.

## **1.2 OBJETIVOS**

### **1.2.1 Objetivo Geral**

O objetivo central deste estudo consistiu em desenvolver e avaliar o desempenho de classificadores baseados em rede neurais artificiais, especificamente rede neural de cápsula e rede convolucional profunda, para a tarefa de classificação de lesões da laringe em imagens de banda estreita obtidas por meio de exame de laringoscopia, utilizando recurso de aumento da base de dados com GAN e programação visual.

### 1.2.2 Objetivos Específicos

Para o alcance do objetivo principal deste estudo, constituem-se como objetivos secundários:

- Desenvolvimento de um classificador baseado em algoritmo de rede neural em cápsula;
- Adoção de técnica de data augmentation para ampliação da base de dados com imagens de tecido da laringe, geradas sinteticamente a partir de uma rede generativa adversária.
- Avaliação do desempenho do classificador baseado em cápsula, comparativamente a outros classificadores baseados em *machine learning* explorados na literatura;
- Avaliação do impacto do recurso de data augmentation com imagens geradas por GAN na performance do classificador baseado em rede de cápsula;
- Desenvolvimento de um classificador baseado em CNN utilizando software de programação visual.

### 1.3 JUSTIFICATIVA

O Carcinoma Espinocelular da Laringe (CECL) é a neoplasia mais frequente na laringe, correspondendo a cerca de 98% de todos os casos de câncer que acometem esse órgão. Esse tipo de tumor maligno é responsável pela maioria dos adoecimentos por câncer da região da cabeça e pescoço.

O diagnóstico tardio da doença implica na adoção de tratamentos mais agressivos, que afetam negativamente a qualidade de vida, bem como eleva as taxas de mortalidade e a necessidade de ressecção da laringe e das pregas vocais.

Uma dificuldade existente para o diagnóstico do CECL em estágio inicial consiste na identificação de lesões menores na mucosa da laringe, uma vez que tais alterações são menos perceptíveis ao olhar do examinador.

A automatização da classificação de imagens de banda estreita da laringe, pode permitir a redução de erros diagnósticos e possibilitar a intervenção precoce do profissional médico, evitando procedimentos que resultem no comprometimento da estrutura anatômica da laringe e contribuindo para a melhora do prognóstico da doença.

Pesquisas relacionadas ao objeto de estudo desta dissertação apontam que a adoção de metodologias baseadas de aprendizado de máquina, como máquinas de vetores de suporte,

árvores de decisão e redes neurais CNN, com o objetivo de auxiliar o diagnóstico de lesões da laringe. Tais pesquisas têm obtido resultados relevantes na identificação de lesões e de carcinoma da laringe. Em estudo conduzido por Araújo *et. al*(2019) (Araújo *et al.*, 2019), os autores alcançaram sensibilidade com o uso de descritores globais baseados em textura e descritores baseados em aprendizagem profunda com CNN. Outro trabalho, realizado por Moccia *et. al* (2017) (Moccia *et al.*, 2017), foram utilizados diversos modelos de classificadores multiclasse baseados em aprendizado de máquina para detectar lesões, alcançando sensibilidade de 98% no melhor desempenho.

Porém, apesar dos resultados reportados com CNN, não foram identificados estudos envolvendo modelos mais novos de redes neurais artificiais para detecção de lesões em imagens da laringe obtidas por NBI, como as redes neurais de cápsula, idealizadas com o propósito de superar limitações do modelo CNN, e GAN, que podem ser utilizadas para gerar dados artificiais para aprimorar o treinamento de classificadores.

Um outro aspecto relevante é que os trabalhos encontrados referem ao desenvolvimento de soluções complexas de *machine learning* que exigem conhecimentos em programação para implementação e utilização das ferramentas desenvolvidas.

Nesse sentido, o estudo e o desenvolvimento de novas técnicas de classificação envolvendo rede de cápsulas e aumento da base de dados com GAN, com o objetivo de facilitar o diagnóstico de lesões em fase inicial tal como é proposto nesta pesquisa, pode beneficiar os indivíduos afetados ao possibilitar a identificação tempestiva de lesões precursoras câncer de laringe. Ainda, o desenvolvimento de uma solução de inteligência artificial baseada em programação visual pode democratizar o acesso de usuários leigos em computação a uma ferramenta de classificação automatizada e de auxílio ao diagnóstico de lesões da laringe.

## 1.4 ESTRUTURA DA DISSERTAÇÃO

A estrutura desta dissertação está dividida em seis partes. Além desta introdução, o presente trabalho está constituído dos seguintes capítulos: fundamentação teórica, materiais e métodos, resultados e discussão, conclusão e trabalhos futuros.

O segundo capítulo, denominado "Fundamentação Teórica", apresenta um levantamento bibliográfico e explanação dos conceitos relativos ao carcinoma espinocelular da laringe, lesões da laringe, processamento digital de imagens, redes neurais artificiais e sua aplicação na tarefa de classificação de imagens médicas.

A seção "Materiais e Métodos" aborda o detalhamento das ferramentas selecionadas para realização do experimento e de cada uma das etapas e procedimentos que constituíram a metodologia adotada para alcançar os objetivos propostos: aquisição das

imagens, os modelos de RNA escolhidos, a técnica de aumento de dados e as métricas utilizadas para avaliação dos resultados.

O quinto capítulo, "Resultados e Discussão", se dedicou a apresentação dos resultados alcançados pelo modelo proposto utilizando tabelas e gráficos, e também à realização de uma análise comparativa utilizando as métricas selecionadas na metodologia e discussão dos resultados obtidos.

No sexto capítulo, "Conclusão", são apresentadas as conclusões extraídas a partir da análise dos resultados discutidos. As conclusões foram elaboradas de forma a responder às questões de pesquisa propostas no início desta pesquisa.

Por fim, na última parte, são apresentadas as limitações e sugestões de trabalhos futuros relacionados ao objeto de estudo da pesquisa realizada, que possam contribuir para a melhoria dos resultados alcançados.

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 CARCINOMA ESPINOCELULAR E LESÕES DA LARINGE

O câncer de laringe é uma neoplasia que representa cerca de 25% dos casos de câncer que acometem a região da cabeça e pescoço (Santiago, 2015).

Esse tipo de tumor ocorre com maior frequência em indivíduos do sexo masculino (Santiago, 2015), e estudos apontam para a existência de relação direta entre elitismo, HPV e tabagismo e maior risco de incidência de câncer de laringe (Menezes *et al.*, 2002) (Abreu *et al.*, 2020).

O CECL ocorre a partir do desenvolvimento anormal das células escamosas que compõem a camada interna do revestimento desse órgão, e pode surgir em qualquer uma de suas subestruturas: glote, subglote e supraglote (Santiago, 2015).

O epitélio escamoso é caracterizado por possuir uma camada de células superficiais, composta por células planas assemelhadas a escamas. A área de epitélio escamoso displásico apresenta maior extensão que a dos epitélios escamoso normal (Soares, 2009).

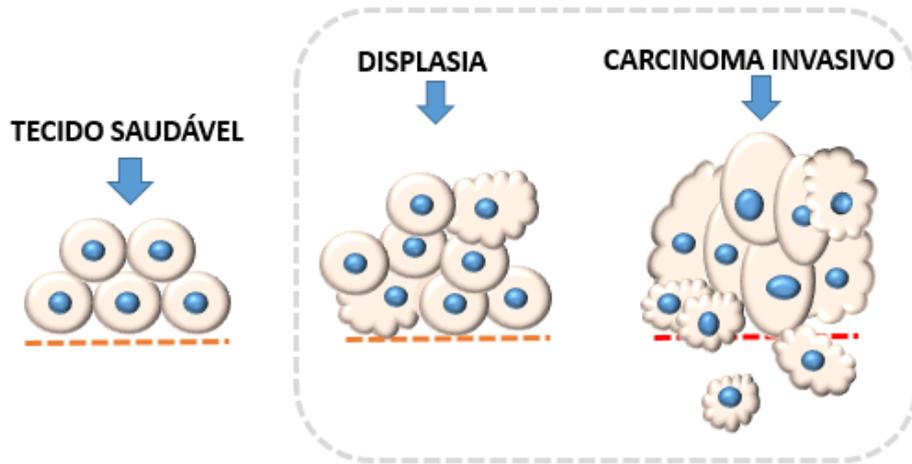
A Figura 2.1 ilustra a diferença entre a organização e forma das células no tecido saudável da mucosa, o tecido com displasia e a mucosa acometida por carcinoma invasivo.

Nos estágios iniciais, o CECL apresenta-se clinicamente assintomático na maioria das vezes, e conforme a doença evolui, manifestam-se sintomas como disfagia e disfonia (Santiago, 2015).

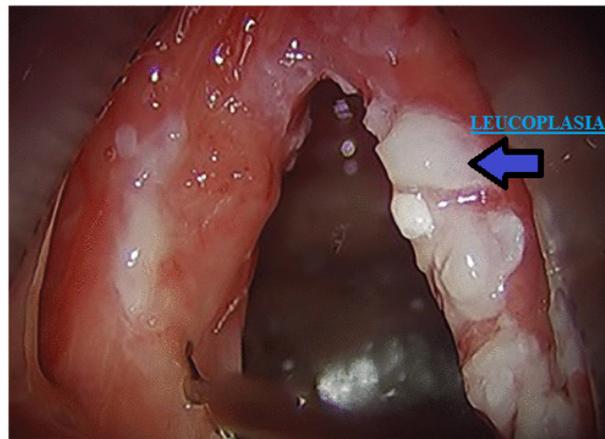
Os principais sinais indicativos de acometimento da laringe por carcinoma espinocelular que se apresentam ao exame de laringoscopia consistem na presença de lesões como leucoplasia e vasos hipertróficos na mucosa.

A leucoplasia da prega vocal laríngea caracteriza-se por uma lesão anormal de mucosa, manifestada por placas brancas epiteliais planas ou espessas ou manchas e que não podem ser definidas como qualquer outra condição (Panwar *et al.*, 2013). A Figura 2.2 apresenta a imagem de uma lesão leucoplásica na mucosa da laringe.

A presença de leucoplasia na laringe não é sinônimo da presença de câncer, menos de 20% das lesões orais são displasias (Gale *et al.*, 2016). Contudo, a leucoplasia da



**Figura 2.1.** Ilustração da diferença entre mucosa saudável, com displasia e com carcinoma invasivo.

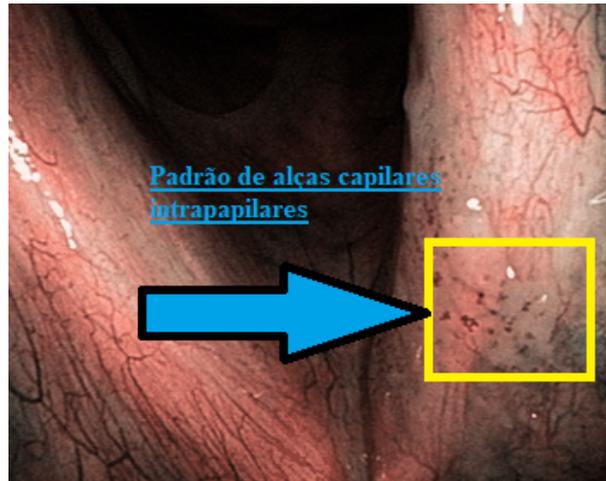


**Figura 2.2.** Leucoplasia na laringe. Adaptado de: (Park *et al.*, 2021).

prega vocal é um motivo de alerta na avaliação clínica, pois tende a se transformar em carcinoma invasivo (Chen *et al.*, 2019).

A presença de vasos hipertróficos aferentes, ramificados em alças vasculares dentro da lesão, também é considerada um indício de área suspeita (Klimza *et al.*, 2019), conforme exemplo da Figura 2.3.

O tecido saudável da mucosa da laringe normalmente apresenta aparência homogênea, sem a presença de lesões ou padrões vasculares em alças intrapapilares. A Figura 2.4 exhibe um exemplo de capturas, obtidas por NBI, de tecido saudável da laringe.



**Figura 2.3.** Padrão de Alças Capilares Intrapapilares. Adaptado de: (Lukes *et al.*, 2013).



**Figura 2.4.** Exemplo de capturas em NBI de tecido saudável da laringe. Adaptado de: (Moccia *et al.*, 2017).

NBI é uma técnica endoscópica considerada moderna, caracterizada pela utilização de filtros especiais que selecionam apenas os comprimentos de onda da luz verde (540 nm) e da luz azul (415 nm), correspondentes aos picos de absorção da hemoglobina. Estes comprimentos de onda penetram nas camadas mais superficiais da mucosa da laringe, realçando a rede capilar e os vasos submucosos, e permitindo melhor identificação de pequenas lesões e conseqüentemente, padrões neoangiogênicos, que passam despercebidas pela endoscopia de luz branca. (Piazza *et al.*, 2012) (Ni *et al.*, 2011).

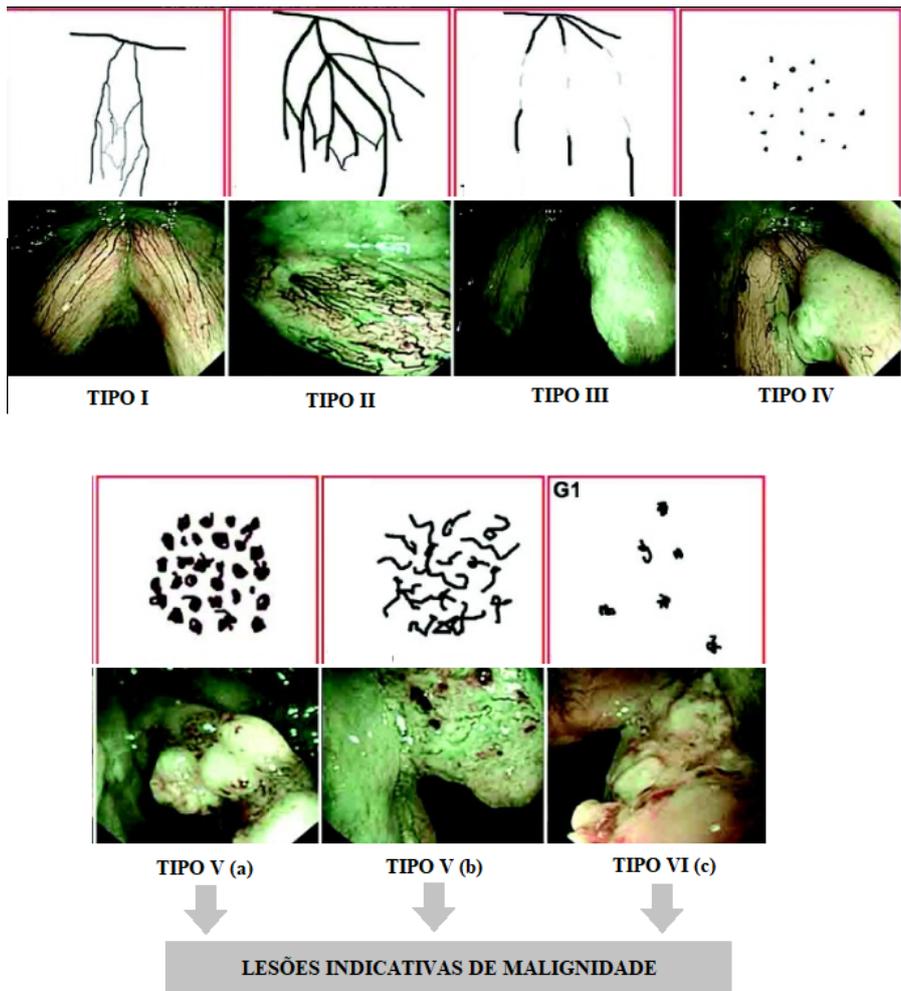
NBI apresenta elevado nível de sensibilidade e especificidade na predição de feridas neoplásicas, mesmo após a realização de procedimentos de radioterapia ou quimiorradiação (Piazza *et al.*, 2010).

Ni *et al.* (2011) (Ni *et al.*, 2011), classifica as alterações das alças capilares intrapapilares visíveis em imagens de banda estreita em cinco tipos (Figura 2.5):

- Tipo I - as alças capilares papilares intraepiteliais são quase invisíveis, e vasos oblíquos e arborescentes de pequeno diâmetro podem ser visualizados com clareza;
- Tipo II - as alças capilares papilares intraepiteliais quase invisíveis, porém vasos oblíquos e arborescentes apresentam diâmetro aumentado;
- Tipo III - mucosa branca com alças capilares papilares intraepiteliais que não podem

ser visualizadas; se a mancha branca for fina, os vasos oblíquos e arborescentes podem ser vistos indistintamente, mas se a mancha branca for espessa, os vasos ficarão obscurecidos;

- Tipo IV - as alças capilares papilares intraepiteliais da mucosa exibem um padrão regular e baixa densidade, com terminais capilares são bifurcados ou ligeiramente dilatados, e as alças capilares papilares intraepiteliais aparecem como pequenas manchas marrom-escuras dispersas com vasos oblíquos e arborescentes pouco visíveis;
- Tipo V - lesões indicativas de malignidade, que podem se apresentar de três maneiras: (a) alças capilares papilares intraepiteliais significativamente dilatadas e de densidade relativamente alta, e parecem ser sólidas ou ter características ocas, acastanhadas, salpicadas e várias formas; (b) a alça capilar papilar intraepitelial é destruída, com seus remanescentes apresentando-se em forma de ramos, e os microvasos são dilatados e alongados; e (c) superfície da lesão recoberta por tecido necrótico, com alças capilares papilares intraepiteliais apresentando-se sob a forma de manchas acastanhadas ou tortuosas de densidade desigual espalhadas de forma irregular sobre a superfície do tumor.



**Figura 2.5.** Classificação dos Tipos de Lesões em Alças Capilares Intrapapilares na Laringe. Adaptado de: (Ni *et al.*, 2011).

## 2.2 PROCESSAMENTO DIGITAL DE IMAGENS

O Processamento Digital de Imagens (PDI) consiste num conjunto de técnicas computacionais que modificam uma imagem de entrada produzindo valores de saída desejados (Gonzalez e Woods, 2008). Conforme Marques Filho e Vieira Neto (1999), o PDI proporciona melhoria nas informações de uma imagem digital para auxiliar na análise pela visão humana ou por computador, de forma automatizada (Filho e Neto, 1999).

Uma imagem monocromática é caracterizada por uma função bidimensional  $f(x, y)$  em que  $x$  e  $y$  são coordenadas no espaço e a amplitude da função  $f$ , para cada par de coordenadas, corresponde à intensidade ou grayscale, conforme exemplo constante na Figura 2.6. A imagem é classificada como digital quando os valores das coordenadas  $x$  e  $y$  e da amplitude de  $f$  assumem valores finitos. Assim, a imagem digital pode ser definida como a representação de uma cena ou imagem analógica sob forma de uma matriz bidimensional. Os menores elementos de uma imagem digital são denominados pixels (Gonzalez *et al.*, 2004).



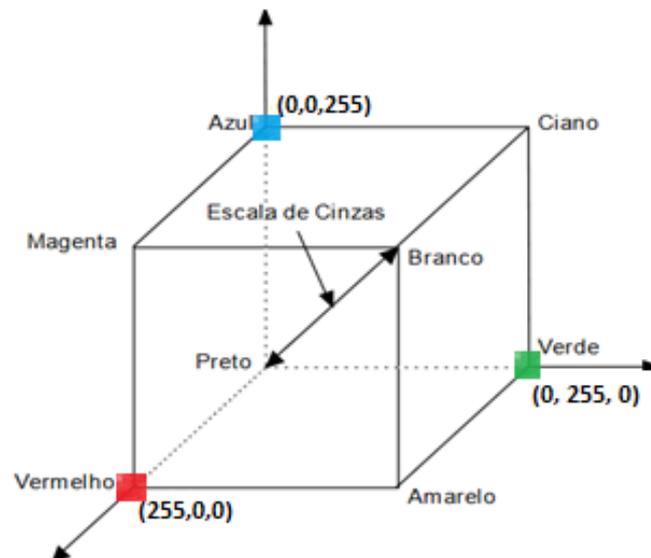
**Figura 2.6.** Imagem digital monocromática representada nas coordenadas X e Y. Adaptado de: (Filho e Neto, 1999).

Imagens coloridas são constituídas pela composição de duas ou mais funções bidimensionais monocromáticas  $f(x, y)$ , em que cada função é chamada de banda ou canal de frequência. Cada pixel da imagem colorida possui a informação relativa à cor de todos canais que a compõem (Filho e Neto, 1999).

As cores são importantes descritores das propriedades de um objeto. Um dos modelos mais utilizados para representação de cores em imagens digitais é o sistema RGB, constituído por três canais, R, G e B, que carregam a informação da intensidade das cores primárias aditivas, respectivamente, vermelho (red), verde (green) e azul (blue).

O espaço de cores RGB é usualmente representado através de um cubo, em que: três vértices constituem as cores primárias, outros três correspondem às cores secundárias, o

ponto localizado na origem representa a cor preta e seu ponto antipodal corresponde à cor branca. O segmento de reta que liga o ponto localizado na origem, referente à cor preta, ao ponto representado pela cor branca é definido como a escala de cinza, conforme demonstrado na Figura 2.7. (Filho e Neto, 1999)



**Figura 2.7.** Representação do Espaço de Cores RGB. Adaptado de: (Filho e Neto, 1999).

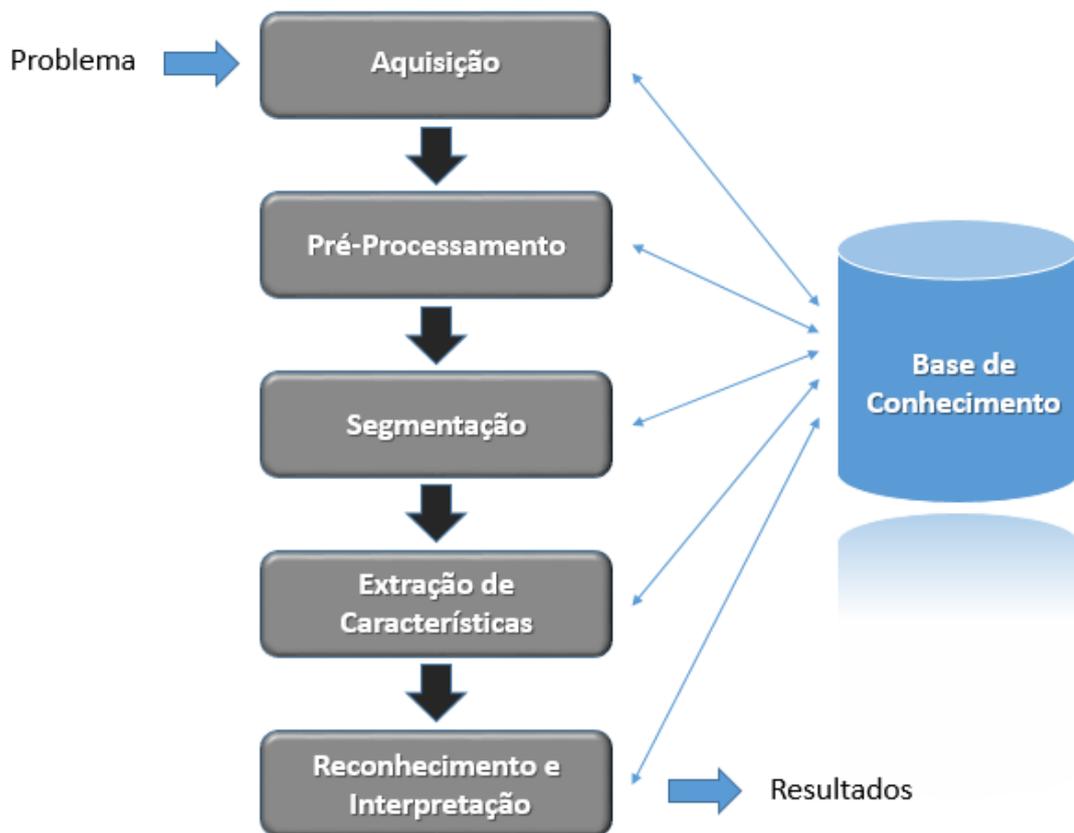
### 2.2.1 Etapas do Processamento Digital de Imagens

Conforme Gonzalez e Woods (2008), é possível dividir PDI em cinco etapas fundamentais (Figura 2.8): aquisição, pré-processamento, segmentação, extração das características, reconhecimento e interpretação.

A primeira etapa do PDI se concentra na obtenção de imagens em formato digital, adequadas para o processamento digital.

Após a aquisição das imagens, é dada sequência à fase de pré-processamento ou melhoramento, em que são realizadas manipulações na imagem de entrada com o objetivo de aprimorar a qualidade desta e preparando-a para as próximas etapas do processamento, de tal forma que a imagem resultante possua melhor qualidade que a original. Assim essa fase é dedicada a promover a redução de imperfeições que possam constar na imagem adquirida com a utilização de técnicas diversas, tais como ajuste de contraste, correção de ruídos e aplicação de filtros.

A segmentação, por sua vez, consiste na tarefa de dividir a imagem em porções menores, permitindo a focalização nos objetos de interesse.



**Figura 2.8.** Etapas do Processamento Digital de Imagens.

A segmentação da imagem é uma etapa relevante em tarefas de PDI, visto que é nesta etapa em que são identificados e reconhecidos os objetos de interesse. De acordo com Gonzales e Woods, os algoritmos utilizados para segmentação de imagens baseiam-se em duas propriedades: similaridade e descontinuidade. O conceito de similaridade se baseia no agrupamento de pixels de uma determinada região da imagem e que possuem semelhanças entre si. Descontinuidade relaciona-se à diferença abrupta em regiões de pixels, que sinaliza a existência de bordas. Existem diferentes métodos de segmentação de imagens, dentre os quais a limiarização de Otsu, k-Means e até mesmo aplicações de redes neurais artificiais projetadas para a tarefa de segmentação.

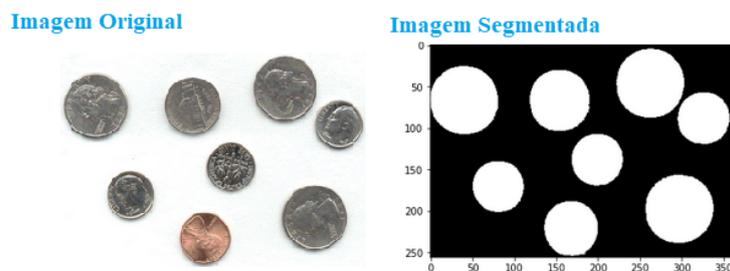
Criado por Nobuyuki Otsu em 1975, o método de segmentação de Otsu é um algoritmo de limiarização que se empenha em determinar, automaticamente, um valor de limiar ideal para segregar os objetos e o fundo da imagem, atribuindo-lhes a cor branca ou preta, tendo por resultado uma imagem de saída binária. A partir da imagem de entrada em tons de cinza, o método calcula seu histograma. Em seguida, são calculados a média e o peso de cada classe. O valor de *threshold* ótimo é aquele que maximiza a variância interclasses. Esse limiar ótimo será selecionado para iterar cada pixel da imagem e modificar sua cor: caso o valor seja inferior ao limiar definido, a cor do pixel é alterada para branca,

caso contrário, o pixel é alterado para a cor preta. O algoritmo de limiarização de Otsu apresenta melhor desempenho em imagens que possuem histograma bimodal, com objetos e fundo bem definidos.

O método K-Means é um dos primeiros algoritmos de clusterização definidos na literatura científica. K-Means é um algoritmo não-hierárquico que se baseia na definição de centróides para um número pré-definido  $k$  de clusters. Primeiramente, o algoritmo distribui de maneira aleatória os pixels entre  $k$  clusters e obtém os centróides através do cálculo da média dos pontos do cluster. Cada pixel da imagem é atribuído a um cluster de forma que seja mínima a distância quadrática média entre o pixel e o centróide mais próximo, e é realizado o cálculo da média do cluster novamente. Esse último passo é iterado até que não haja nenhuma troca de pixels entre clusters, obtendo-se a imagem segmentada.

Uma abordagem mais recente para a tarefa de segmentação de imagens, inclusive imagens médicas, consiste na adoção de modelos baseados em aprendizado de máquina (Hesamian *et al.*, 2019). FCN, U-Net, SegNet e SegCaps constituem alguns modelos de redes neurais artificiais utilizados na segmentação automática de imagens. SegCaps é um método proposto por LaLonde (LaLonde *et al.*, 2021) que utiliza redes de cápsula para segmentação de objetos, demonstrando eficácia na segmentação de imagens biomédicas. (Kosgiker *et al.*, 2021)

A Figura 2.9 apresenta um exemplo de segmentação de imagem.



**Figura 2.9.** Exemplo de segmentação de imagem.

A partir das imagens obtidas pelo processo de segmentação, são extraídas características determinantes, através de descritores que permitam a diferenciação dos objetos com um nível satisfatório de precisão.

Na etapa de extração de características, atributos importantes das imagens são transformados em dados quantitativos de forma e utilizados na etapa subsequente, de reconhecimento dos objetos de interesse. Atributos como espaço de cor, textura e relação de posição entre dois ou mais objetos que compõe a imagem, são capturados da imagem digital nesta etapa.

A última etapa do PDI consiste em no reconhecimento e interpretação dos objetos

segmentados. Nessa etapa é realizada a classificação, com a atribuição de rótulos aos objetos identificados na imagem, conforme suas características, e a interpretação do significado dos resultados obtidos. Algoritmos de aprendizagem de máquina tem se tornado um método bastante utilizado para promover reconhecimento de imagens e objetos, com excelentes resultados.

## 2.3 MINERAÇÃO DE DADOS E CLASSIFICAÇÃO DE IMAGENS

### 2.3.1 Mineração de Dados

Conforme Fayyad *et. al* (1996), mineração de dados consiste na aplicação de algoritmos específicos para extrair padrões de dados (Fayyad *et al.*, 1996). Algumas das principais tarefas de mineração de dados são:

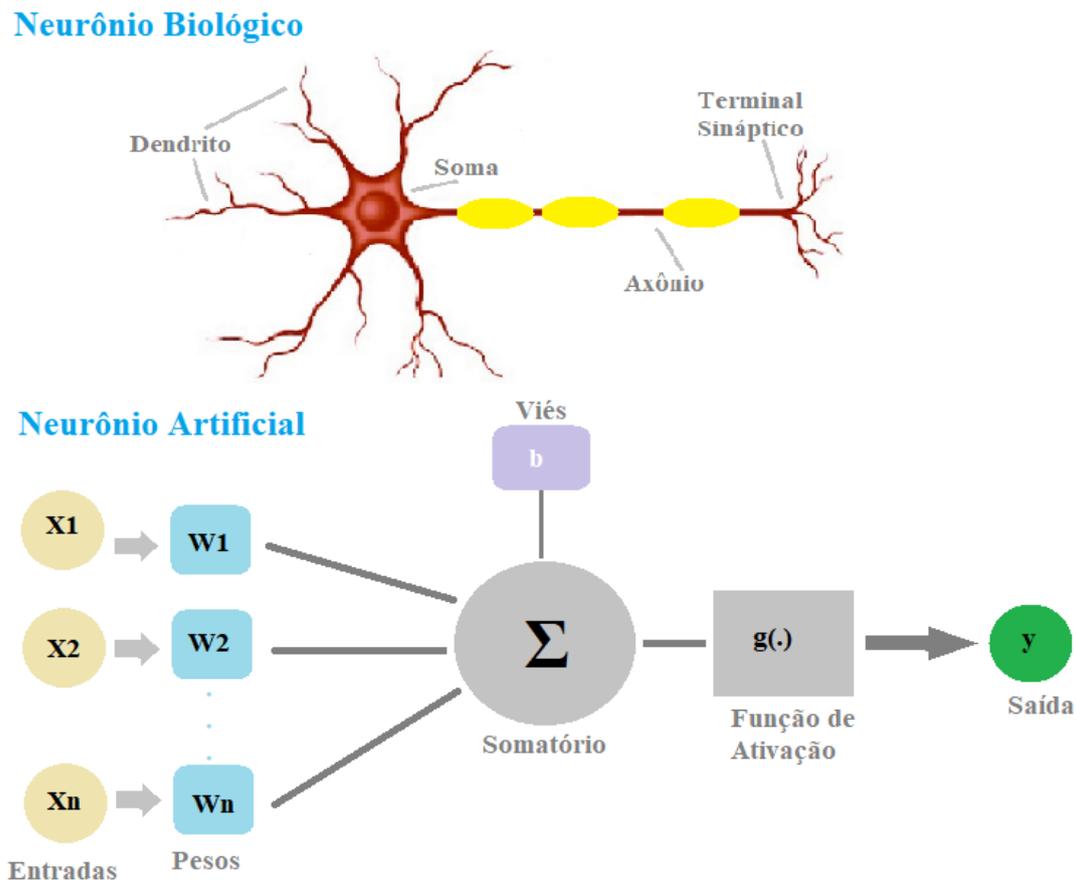
- Descoberta por Associação, que consiste na busca relações ou padrões frequentes entre dados;
- Classificação, que busca pela identificação de conjuntos de dados em que os elementos possuem características semelhantes, rotulando-as em classes;
- Regressão, em que-se busca por uma função, linear ou não, capaz de mapear os registros de um conjunto de dados, limitando-se a atributos numéricos;
- Clusterização, consiste no segmentação automática da base de dados em subconjuntos de itens que compartilham atributos comuns, abstendo-se, no entanto, rotulá-los.

Há uma vasta gama de algoritmos de aprendizado de máquina para mineração de dados. A seleção da técnica adequada é fundamental para que um modelo obtenha desempenho satisfatório na tarefa proposta. Uddin *et. al* (2019) (Uddin *et al.*, 2019), em estudo comparativo dos diferentes algoritmos de aprendizado de máquina supervisionados aplicados em mineração de dados para predição de doenças, identificou maior prevalência da utilização do algoritmo Máquina de Vetores de Suporte (do inglês *Support Vector Machine*) (SVM), sendo, porém o algoritmo Floresta Aleatória (do inglês *Random Forest*) (RF) aquele que apresentou melhor precisão comparativa.

Entretanto, especificamente em relação à classificação de imagens médicas, Yadav *et al.* (Yadav e Jadhav, 2019) aponta para a aprendizagem por transferência baseada em redes neurais profundas como melhor método entre os métodos tradicionais, porque podem aprender e selecionar recursos de forma automática e eficaz, destacando-se a arquitetura de rede de cápsula.

### 2.3.2 Redes Neurais Artificiais e Classificação de Imagens Médicas

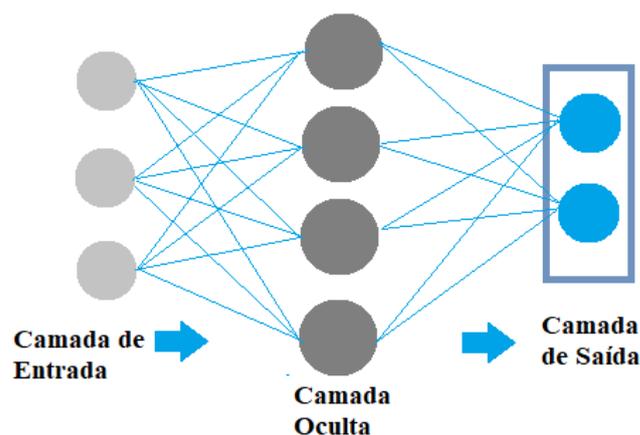
O desenvolvimento de uma RNA baseia-se na simulação computacional do funcionamento do cérebro humano, inspirando-se no processo de aprendizagem e transmissão de informações pelos neurônios biológicos. A Figura 2.10 descreve a estrutura de um neurônio artificial e um neurônio biológico.



**Figura 2.10.** Estrutura de um neurônio artificial. Adaptado de (Borges *et al.*, 2015)

Uma rede neural artificial baseada no conceito de *backpropagation* introduzido por Hinton, Rumelhart e Williams (Rumelhart *et al.*, 1986) possui como estrutura típica uma camada de entrada, uma camada de saída e pelo menos uma camada "oculta" intermediária, em que cada camada está totalmente conectada à seguinte. Um algoritmo de retropropagação fundamenta-se na utilização de gradientes, e consiste na minimização dos erros de propagação, partindo do erro da saída e retrocedendo esse erro em direção à entrada, atualizando os pesos entre as conexões dos neurônios artificiais. Na Figura 2.11, é representada a estrutura básica das camadas que constituem uma rede neural artificial.

Os modelos de aprendizagem profunda se difundiram no meio acadêmico a partir de 2006 com a publicação do artigo "Reducing the Dimensionality of Data with Neural



**Figura 2.11.** Estrutura de uma rede neural artificial.

Networks”, em que G.Hinton e R. Salakhutdinov demonstraram a capacidade de treinar previamente uma rede neural de múltiplas camadas, uma camada de cada vez (Hinton e Salakhutdinov, 2006). Desde então, diversas arquiteturas de redes neurais profundas foram desenvolvidas e aplicadas com objetivo de aperfeiçoar e automatizar a realização de atividades humanas.

Redes CNN são uma categoria de redes neurais artificiais baseadas em processos biológicos do sistema visual (Matsugu *et al.*, 2003).

Consistem em um conjunto de camadas distintas de convolução e subamostragem empilhadas, que recebem um vetor e, por meio de funções de ativação, o transformam em um valor escalar. Por meio desse mecanismo, a CNN pode reconhecer características mais específicas da imagem, como bordas, na primeira camada, e partes completas do objeto na última camada (Aloysius e Geetha, 2018).

A camada de convolução é constituída por múltiplos neurônios, que aplicam um filtro em determinada região da imagem processada. Para cada neurônio que compõe a rede, há uma conexão com um subgrupo de pixels da camada anterior, sendo que cada uma dessas conexões é mapeada por meio da atribuição de pesos. Por sua vez, a camada de pooling implementa uma função não-linear de subamostragem para diminuir a dimensionalidade e capturar pequenas invariâncias (Haffeman, 2014).

Em que pese CNN alcançarem bom desempenho na classificação de imagens, sua arquitetura apresenta uma limitação significativa na classificação de objetos que compõem uma imagem, devido ao seu procedimento roteamento, uma vez que, rotear as camadas através de *pooling*, informações como localização e posicionamento de objetos componentes são desconsideradas (Santos *et al.*, 2020).

A Figura 2.12 descreve a estrutura básica de uma CNN.

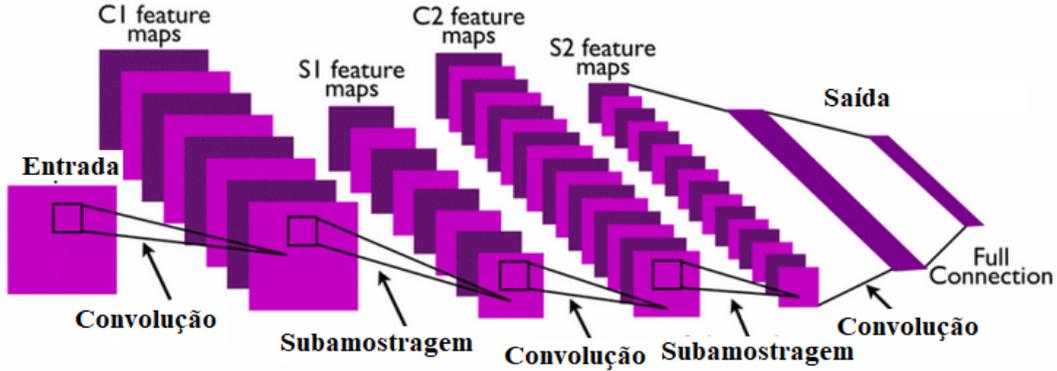


Figura 2.12. Estrutura de uma rede CNN em dois estágios (LeCun *et al.*, 2010).

A Rede Neural em Cápsula (CapsNet), proposta pela primeira vez por Sabour *et al.* em 2017 (Sabour *et al.*, 2017), consiste em uma arquitetura de rede neural profunda baseada em aprendizado profundo projetada para ser capaz de detectar objetos dispostos em uma imagem e suas propriedades, bem como conservar suas relações hierárquicas, na capacidade de conservar a informação sobre a relação dos objetos e da imagem como um todo durante o procedimento de classificação. Assim, uma vantagem das redes em cápsulas em comparação às redes convolucionais tradicionais reside na capacidade de conservar a informação sobre a relação dos objetos e da imagem como um todo durante o procedimento de classificação.

Em decorrências dessas características, as redes de cápsula têm sido aplicadas no processamento de imagens médicas, obtendo desempenho próximo ou superior ao apurado por outros tipos de classificadores (Afshar *et al.*, 2020) (Yadav e Jadhav, 2019).

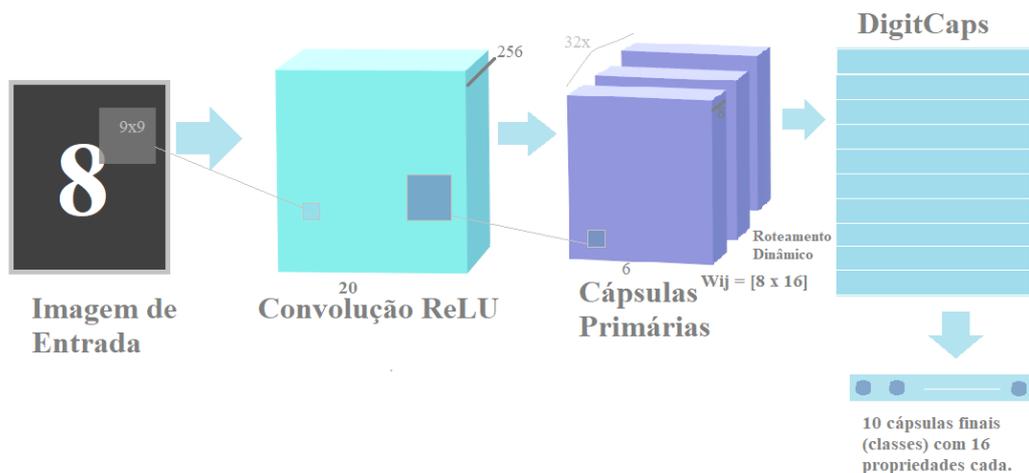
O vetor de saída de cada cápsula é calculado utilizando uma função de ativação chamada "squashing", descrita em exemplo na Equação 2.1, que ajusta os valores de forma que o módulo do vetor fique posicionado em um intervalo entre 0 e 1, preservando as proporções dos parâmetros e permitindo que seja representada a probabilidade da presença de um determinado objeto na imagem.

$$V_j = \frac{\|S_j\|^2}{1 + \|S_j\|^2} \cdot \frac{S_j}{\|S_j\|} \quad (2.1)$$

onde  $V_j$  é o vetor de saída da cápsula  $j$  e  $S_j$  é a soma dos vetores de entrada.

A comunicação entre as cápsulas é realizada a partir de um algoritmo de roteamento dinâmico, que considera o grau de concordância entre as predições das cápsulas comunicantes, permitindo a captura das relações entre os elementos do objeto e o todo.

A estrutura que compõe a arquitetura da rede neural em cápsula proposta por Sabour pode ser visualizada a partir do esquema descrito na Figura 2.13.



**Figura 2.13.** Esquema de arquitetura de rede em cápsula. Adaptado de: (Sabour *et al.*, 2017).

### 2.3.3 Data Augmentation

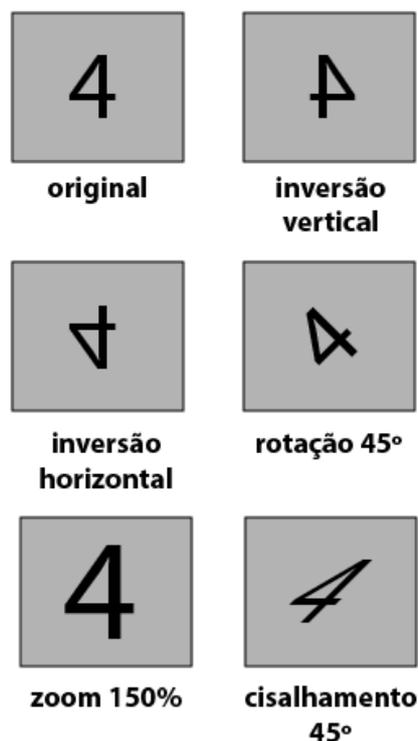
O desempenho de algoritmos de RNA baseados em aprendizagem profunda pode ser melhorado na tarefa de classificação de imagens a partir de técnicas que permitam ampliar o conjunto de imagens que serão utilizados no treinamento da rede.

O emprego de técnicas de processamento de imagens com o objetivo de ampliar a base de dados disponível, frequente em aplicações de *machine learning*, é denominado *data augmentation*.

No contexto de classificação de imagens, transformações geométricas como rotação, translação, inversão da imagem nos eixos horizontal ou vertical, ampliação e cisalhamento, transformações das características de cores, implementação de filtros e mesclagem de imagens, constituem mecanismos classicamente utilizados como recursos de aumento de dados em algoritmos de aprendizado profundo, conforme exemplificado na Figura 2.14.

Outro método de *data augmentation* que recentemente tem sido utilizado no contexto de classificação de imagens com aprendizado de máquina consiste no aprimoramento com GAN.

Redes GAN consistem em um tipo de rede neural capaz gerar amostras artificiais com base em suposições sobre a distribuição dos dados. É composta em um modelo “discriminador”, que aprende uma distribuição de probabilidade condicional, classificando um dado exemplo  $x$  diretamente em um rótulo  $y$ ; e um modelo “gerador”, que aprende a distribuição de probabilidade conjunta  $p(x, y)$ , onde  $x$  representa os dados de entrada e  $y$  o rótulo que se deseja classificar (Oliveira, 2019).



**Figura 2.14.** Data Augmentation com Transformações Geométricas

A utilização de redes GAN se popularizou nos últimos anos, com aplicação em diversas áreas do conhecimento, pela capacidade de criar amostras sintéticas de imagens, áudios e vídeos, com alto grau de semelhança com os modelos reais, a partir de um conjunto de amostras originais.

Estudos apontam que a ampliação das bases de dados com imagens sintéticas geradas por GAN aumenta a variabilidade dos dados, aperfeiçoando o treinamento dos classificadores deep learning (dos Anjos *et al.*, 2019) (Frid-Adar *et al.*, 2018) (Sandfort *et al.*, 2019) (Zhao *et al.*, 2018).

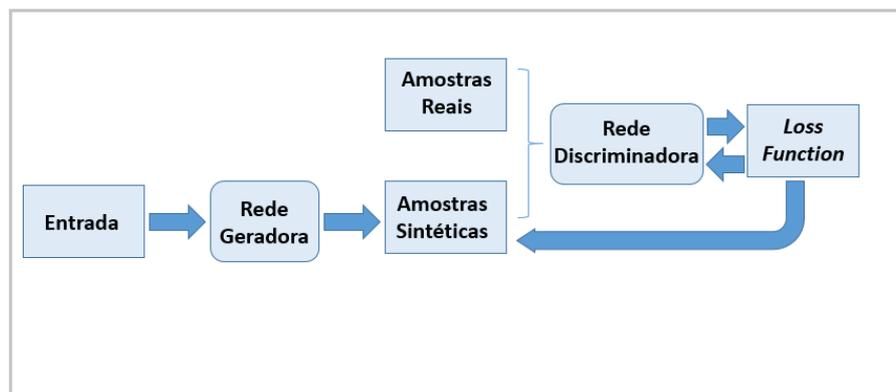
Anjos *et al.* (2019) (dos Anjos *et al.*, 2019), avaliou a qualidade de imagens sintéticas de nódulos pulmonares geradas por GANs com o objetivo de aumentar a base de dados e verificou um aumento de 24,11% na área da Curva Característica de Operação do Receptor (ROC) em relação ao conjunto com aumento de base clássico usando transformações de rotação, inversão e translação.

Um estudo conduzido por Frid-Adar *et al.* (2018) (Frid-Adar *et al.*, 2018) apresentou um método que usa a geração de imagens médicas sintéticas para aumento de dados para aumentar o desempenho de classificadores CNN em um problema médico com dados limitados. Os autores demonstraram que a referida técnica aplicada em uma tarefa de classificação de lesão hepática, alcançou uma melhoria de aproximadamente 7% nas métricas de sensibilidade e especificidade usando aumento sintético com GAN, comparado

com o método clássico de aumento de dados.

A geração de imagens sintéticas a partir de GAN tem se mostrado uma solução eficaz para sanar o problema da quantidade reduzida de exames que compõem os repositórios de imagens médicas (Galbusera *et al.*, 2018).

A Figura 2.15 demonstra a estrutura básica de uma GAN.



**Figura 2.15.** Estrutura de uma rede GAN.

### 2.3.4 Ferramentas de Mineração de Dados e Análise de Imagens

Existem diversas ferramentas disponíveis para facilitar a manipulação e a execução de tarefas de mineração de dados. Dentre as principais ferramentas com licença de uso livre, estão incluídas:

- KNIME
- WEKA
- R Software
- Orange Data Mining

#### 2.3.4.1 Software ORANGE

Orange (<http://orangedatamining.com>) é um software livre distribuído sob a licença General Public License, desenvolvido pelo laboratório de Inteligência Artificial da Faculdade de Computação e Ciência da Informação da Universidade de Ljubljana, situada na Eslovênia.

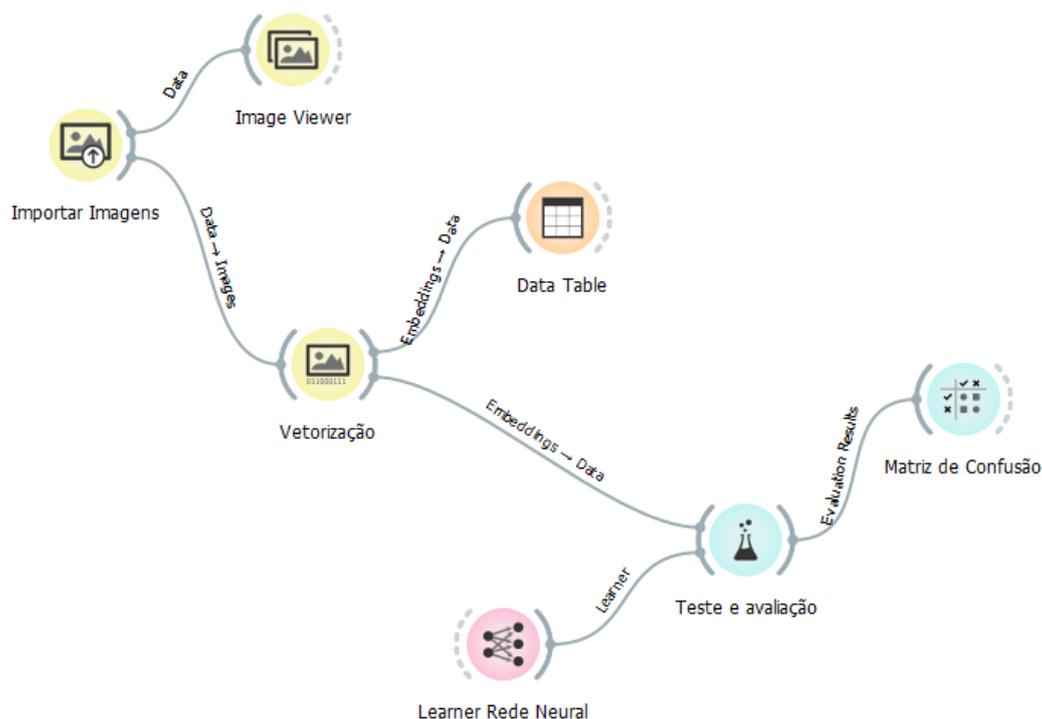
Orange consiste em um sistema baseado em componentes, constituído por objetos e rotinas desenvolvidos em linguagem C++ e contempla diversos algoritmos de aprendizado

de máquina e mineração de dados. Além disso, possui uma interface gráfica denominada Orange Canvas, que permite interligar, visualmente, objetos bem como construir fluxos de processamento para desenvolver modelos de classificação baseados em diferentes algoritmos de aprendizado de máquina, inclusive Árvores de Decisão, SVM e RNA (Demšar *et al.*, 2013).

A ferramenta Orange possui o componente *Image Analytics*, uma extensão gratuita de código aberto para análise de imagens utilizando aprendizado de máquina não supervisionado. Essa solução se baseia na estrutura de mineração de dados de programação visual Orange, aliando a capacidade de construir fluxos de trabalho, desenvolver modelos de dados e permitir visualizações interativas dos dados.

A análise de imagem moderna é bem suportada em ambientes de programação, como aqueles construídos em Python e aprimorada com bibliotecas para aprendizado profundo, como TensorFlow, PyTorch e Keras. O Orange visa complementar essas ferramentas, fornecendo um ambiente acessível e interativo que ainda oferece um alto grau de funcionalidade e pode se adaptar a necessidades específicas por meio de programação visual e construção de problemas (Godec *et al.*, 2019).

A Figura 2.16 exibe um fluxo de análise de imagens utilizando o software Orange:



**Figura 2.16.** Modelo de análise de imagens implementado no software Orange.

### 2.3.4.2 Google Colab

A ferramenta Google Colab consiste em uma plataforma de notebooks Jupyter executada em nuvem. Para este estudo, optou-se pela utilização do Tensorflow junto à biblioteca Keras por ser uma tecnologia acessível, de fácil utilização e bastante popular em comunidades de desenvolvedores de *machine learning*, recebendo constantemente suporte e contribuições. A biblioteca Keras foi utilizada em backend, implementada junto ao Tensorflow, no ambiente Google Colab.

O Tensorflow foi desenvolvido pelo Google Braine e disponibilizado em código aberto em 2015. Consiste em um *framework* para Python, Javascript, C/C++ e Go, que permite o processamento de dados para aprendizado de máquina por meio de um modelo baseado em fluxo de dados Python (Roesler *et al.*, 2018).

O framework Tensorflow suporta treinamento em larga escala e usa eficientemente centenas de servidores poderosos para treinamento rápido e executa modelos treinados para produção de inferência em várias plataformas, desde grandes clusters distribuídos em um datacenter até para executar localmente em dispositivos móveis. Ao mesmo tempo, é flexível o suficiente para oferecer suporte à experimentação e pesquisa de novos modelos de aprendizado de máquina e melhorias no nível do sistema (Abadi *et al.*, 2016).

Keras é uma Interface de Programação de Aplicações, do inglês *Application Programming Interface* (API), de alto nível baseada em Python, desenvolvida com foco na construção e treinamento de modelos de aprendizagem de máquina (Chollet, 2015).

## 2.4 TRABALHOS RELACIONADOS

Com o objetivo de mapear trabalhos correlatos ao objeto de estudo desta dissertação, foi realizado um levantamento de pesquisas relacionadas nas bases de dados Scopus, Science Direct e Web of Science.

A fim de identificar estudos envolvendo a aplicação de métodos de aprendizado de máquina no diagnóstico de cancer laríngeo em imagens obtidas por NBI, foi realizada a pesquisa através do seguinte termo de busca: "(("NBI"OR "narrow band imaging") AND ( "larynx"OR "laryngeal") AND ("cancer"OR "carcinoma") AND ("machine learning"OR "deep learning"))).

A busca na base Scopus retornou 4 resultados, exibidos no Quadro 2.2.

A pesquisa realizada na base Science Direct retornou 17 resultados, considerando artigos de pesquisa. Dos 17 estudos publicados, 15 foram excluídos após leitura do resumo, por não abordar estudo relativo à classificação de câncer de laringe em imagens NBI. Assim, restaram 2 publicações relacionadas ao tema da pesquisa, descritas no Quadro

### 2.3.

Na base Web of Science, a inserção do termo de busca retornou 3 resultados distintos, visto que um dos artigos, conduzido por Esmaeili *et. al* (2020) (Esmaeili *et al.*, 2020), já constava na base Scopus. O resumo dos estudos consta no Quadro 2.4.

Portanto, a busca nas três bases resultou, no total, em 9 estudos distintos sobre a adoção de modelos de *machine learning* para auxiliar o diagnóstico de carcinoma da laringe a partir de imagens obtidas com NBI, publicados no período de 2017 a 2022, conforme demonstrado na Tabela 1.

**Tabela 2.1.** Quantidade de Artigos Relacionados Publicados

Ano	Publicações			
	Scopus	Science Direct	Web of Science	Total por Ano
2017	1	0	0	1
2019	1	1	0	2
2020	2	0	0	2
2021	0	0	1	1
2022	0	1	2	3
<b>Total</b>	4	2	3	9

O levantamento realizado indicou que os tipos de classificadores que apareceram com maior frequência na metodologia dos estudos foram CNN e SVM:

- 6 utilizaram modelos de Inteligência Artificial (IA) baseados em arquitetura CNN;
- 3 utilizaram para classificação SVM;
- 2 publicações mencionaram RF;
- 2 estudos utilizaram o método k-Nearest Neighbours (kNN).

A melhor sensibilidade(98%) reportada nos trabalhos relacionados a classificação multiclasse de imagens foi obtida pelos estudos conduzidos por Araujo *et. al*(2019), combinando escritores globais baseados em textura e descritores baseados em aprendizagem profunda CNN a um classificador SVM.

**Quadro 2.1.** Trabalhos Relacionados - Base *Scopus*

Autor(es)	Resumo
(Moccia <i>et al.</i> , 2017)	Os autores desenvolveram classificadores multiclases para detecção de CECL a partir de imagens obtidas por NBI baseados em 4 diferentes tipos de algoritmo: k-nearest neighbors; algoritmo de Naive Bayes; RF e SVM. O melhor desempenho alcançado, considerando-se a <b>mediana dos valores de sensibilidade, foi de 93%</b> Intervalo Interquartil (IQR) = 6%]. Adicionalmente, os autores excluíram as imagens de baixa confiança e o valor de sensibilidade mediano alcançado aumentou para 98% (IQR = 5%).
(Inaba <i>et al.</i> , 2020)	Utilizaram RetinaNet para detecção de câncer superficial da laringofaringe em imagens da mucosa laringofaríngea obtidas a partir de imagens de banda estreita. Cada conjunto de dados independente compreendeu 400 de câncer e 800 imagens normais da mucosa. Nos resultados obtidos para os conjuntos de dados de validação, contemplando 100 casos de câncer superficial da laringofaringe, o tamanho médio do tumor foi de 13,2 mm; tipos planos/elevados/deprimidos foram encontrados em 77/21/2 casos, alcançando <b>sensibilidade, especificidade e precisão de 95,5%, 98,4% e 97,3%</b> , respectivamente, <b>na classificação binária de câncer laringofaríngeo.</b>
(Esmaeili <i>et al.</i> , 2020)	Os autores avaliaram e compararam o desempenho de uma abordagem manual e automática para classificação de lesões laríngeas com base em padrões vasculares em imagens obtidas pela combinação de Endoscopia de Contato (EC) e NBI. Na abordagem manual, seis observadores avaliaram visualmente uma série de imagens EC-NBI pertencentes a um paciente e, em seguida, classificaram o paciente como benigno ou maligno. Para a classificação automática, um <b>algoritmo baseado na caracterização do nível de desordem da embarcação em combinação com quatro classificadores supervisionados</b> foi usado para classificar as imagens EC-NBI. Os resultados <b>mostraram que a avaliação subjetiva da abordagem manual pode ser reduzida usando uma abordagem baseada em computador.</b>
(Araújo <i>et al.</i> , 2019)	Investigaram uma abordagem baseada em aprendizagem para o diagnóstico de CECL em estágio inicial e comparar <b>o uso de (i) descritores globais baseados em textura, como padrões binários locais, e (ii) descritores baseados em aprendizagem profunda (CNN).</b> Essas características, extraídas de imagens endoscópicas de banda estreita da laringe, foram classificadas com <b>SVM</b> para discriminar tecidos CECL saudáveis, pré-cancerosos e em estágio inicial. Quando testado em um conjunto de dados de benchmark, <b>uma sensibilidade de classificação mediana de 98%</b> foi obtida com a melhor combinação de recursos, superando o estado da arte (sensibilidade = 95%).

Fonte: elaboração própria.

**Quadro 2.2.** Trabalhos Relacionados - Base *Science Direct*

Autor(es)	Resumo
(Xiong <i>et al.</i> , 2019)	Foi desenvolvido um sistema de diagnóstico baseado em <i>Deep CNN</i> foi construído e treinado usando 13.721 imagens laringoscópicas de LCA, lesões laringeas pré-cancerosas, tumores laringeos benignos (BLT) e tecidos normais (NORM) de 2 hospitais terciários na China. <b>O sistema DCCN alcançou uma sensibilidade de 0,731, uma especificidade de 0,922, uma AUC de 0,922 e uma precisão geral de 0,867 para detectar lesões laringeas cancerosas e pré-cancerosas entre todas as lesões e tecidos normais.</b> Quando comparada a especialistas humanos em um conjunto de teste independente, o desempenho do DCCN na detecção das lesões alcançou uma sensibilidade de 0,720, uma especificidade de 0,948, uma AUC de 0,953 e uma precisão geral de 0,897.
(Kwon <i>et al.</i> , 2022)	Para o diagnóstico de câncer glótico, modelos CNN baseados em aprendizagem profunda foram estabelecidos e classificados usando <b>imagem laringea e dados de voz</b> . A acurácia da classificação foi obtida por meio da aprendizagem do conjunto de árvores de decisão usando probabilidade por meio do algoritmo de classificação CNN. Nesse processo, foi utilizado o método de árvore de classificação e regressão (CART). Em seguida, comparamos a precisão da classificação do aprendizado de <b>conjunto de árvore de decisão com classificadores individuais CNN</b> , fundindo a imagem laringea com o classificador de árvore de decisão de voz. Foi utilizado o aprendizado por ensemble de árvore de decisão da imagem laringea e da voz, e a precisão da classificação foi melhorada pela integração dos dados da imagem laringea e da voz da mesma pessoa. A precisão da classificação foi de 87,88% e 89,06% para a imagem laringea individualizada e o modelo de árvore de decisão de voz, respectivamente, e a fusão dos resultados da imagem laringea e da árvore de decisão de voz representou uma <b>precisão de classificação de 95,31%</b> .

Fonte: elaboração própria.

**Quadro 2.3.** Trabalhos Relacionados - Base *Web of Science*

Autor(es)	Resumo
(Azam <i>et al.</i> , 2022a)	Os autores desenvolveram uma nova aplicação de inteligência artificial para <b>detecção em tempo real de CECL em videolaringoscopias com luz branca e NBI</b> , baseada em modelo <b>You-Only-Look-Once (YOLO)</b> de aprendizagem profunda CNN. Vídeos de CECL foram coletados retrospectivamente de videoendoscopias transnasais realizadas em consultório e endoscopias rígidas intraoperatórias, e frames foram extraídos para treinamento, validação e teste. Os modelos YOLO foram treinados após distribuição aleatória de imagens em um conjunto de treinamento (82,6%), conjunto de validação (8,2%) e conjunto de teste (9,2%). Entre os vários modelos, o algoritmo ensemble (YOLOv5s com YOLOv5m-TTA) obteve os melhores <b>resultados de detecção de CECL</b> , com métricas de desempenho a par dos resultados relatados por outros modelos de detecção de última geração: <b>0,66 de precisão (valor preditivo positivo), 0,62 de recall (sensibilidade) e 0,63 de precisão média.</b>
(He <i>et al.</i> , 2021)	Um conjunto de 3.458 imagens patológicas foi obtido de 1.228 pacientes submetidos à endoscopia assistida por IA do trato aerodigestivo superior para procurar lesões laringeas suspeitas que foram biopsiadas e examinadas histologicamente. As imagens de patologia foram divididas aleatoriamente em um conjunto de dados de treinamento, um conjunto de dados de validação e um conjunto de dados de teste. O algoritmo usado foi o <b>Inception V3</b> e a <b>AUC foi de 0,994 para o conjunto de dados de validação e 0,981</b> para o conjunto de dados de teste.
(Azam <i>et al.</i> , 2022b)	O estudo teve como objetivo desenvolver uma CNN para segmentação semântica automática de câncer do trato aerodigestivo superior em imagens endoscópicas. Um conjunto de dados de luz branca e quadros de vídeo NBI de CECL foi coletado e anotado manualmente. Um novo <b>modelo de segmentação Deep Learning, SEGMENT</b> , foi projetado tendo por base a arquitetura CNN DeepLabV3+, modificada usando o Xception como backbone e incorporando recursos de conjunto de outras CNN. No estudo, realizado com 219 pacientes e 683 videoframes de CECL, SEGMENT <b>obteve valores medianos: 0,68 interseção sobre união, 0,81 coeficiente de similaridade de dados, 0,95 sensibilidade, 0,78 precisão, 0,97 exatidão.</b>

Fonte: elaboração própria.

## 2.5 CONSIDERAÇÕES FINAIS

A identificação automatizada de lesões na laringe com RNA pode auxiliar a detecção precoce do desenvolvimento de CECL, beneficiando o prognóstico e o tratamento dos pacientes acometidos por essa doença.

Existem diversos modelos e arquiteturas de RNA utilizados, dentre os quais se destaca a CNN. Entretanto, essa arquitetura apresenta limitações para identificar a relação entre as partes e o todo de um objeto. As redes baseadas em cápsulas foram idealizadas com o objetivo de superar essa limitação utilizando cápsulas e algoritmo de roteamento dinâmico por concordância.

Técnicas de *data augmentation* são frequentemente utilizadas para melhorar os resultados obtidos por classificadores baseados em aprendizado de máquina. A ampliação da base de dados com GAN é uma dessas técnicas, e se baseia em aumentar o conjunto de dados disponíveis a partir da criação de amostras artificiais.

Embora existam diversas pesquisas explorando a classificação de tumores na laringe a partir de CNN, há poucos estudos relacionados à utilização de classificadores baseados em redes de cápsula, combinado à técnica de *data augmentation* por GAN, para identificação de câncer a partir de imagens do tecido laríngeo obtidas por NBI.

### 3 MATERIAIS E MÉTODOS

A metodologia deste trabalho consistiu das seguintes etapas, descritas na Figura 3.1: aquisição do dataset contendo imagens NBI da mucosa laríngea; processamento digital e implementação dos classificadores baseados em RNA especificamente modelos CNN e CapsNet; ampliação do *dataset* com uma rede neural do tipo GAN; classificação das imagens do dataset ampliado com o classificador baseado em rede de cápsulas; e por fim, avaliação do desempenho obtido por cada um dos algoritmos.

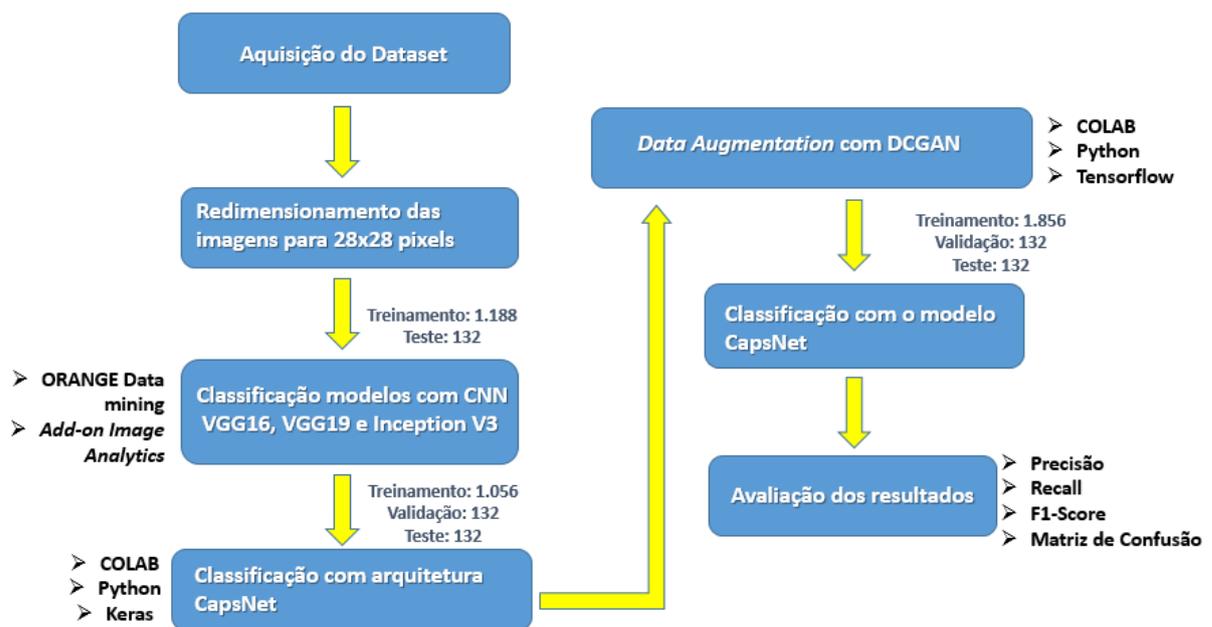


Figura 3.1. Fluxograma com etapas adotadas na metodologia.

Neste estudo foi utilizado o software Orange Data Mining, versão 2.27, por se tratar de ferramenta de livre acesso amplamente explorada no ambiente acadêmico, gratuita, com interface intuitiva e de fácil utilização pelo usuário. Além disso, a opção pela ferramenta se deu pelo fato desta oferecer suporte para utilização de redes neurais artificiais pré-definidas (Inception V3, VGG16, VGG19), através do complemento *Image Analytics*. Para realizar o treinamento das redes neurais a partir do software Orange foi utilizado um computador com processador Intel Core i7, 2.40 GHz e 2 núcleos, com 8 Gigabytes

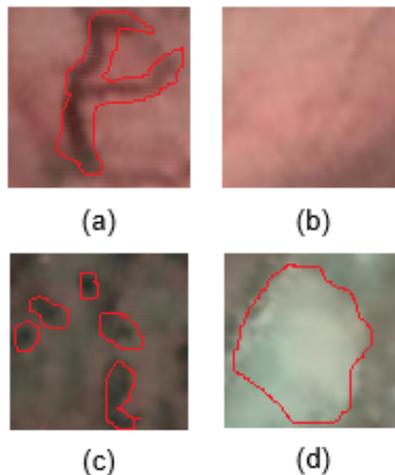
de memória RAM.

Para a implementação das arquiteturas referentes à geração das imagens artificiais com GAN e ao classificador CapsNet em linguagem python foi utilizado um notebook Jupyter executado em nuvem, usando Unidade Gráfica de Processamento (do inglês *Graphical Processing Unit*) (GPU) Tesla V100, com 16GB de RAM, disponibilizada na plataforma Colab para aquisição mediante pagamento, sob a forma de tempo de utilização de unidades computacionais.

### 3.1 AQUISIÇÃO DO BANCO DE IMAGENS

As imagens foram adquiridas a partir de um dataset *open access* composto de 1.320 imagens de tecidos laríngeos saudáveis e cancerosos em estágio inicial, obtidas por Moccia *et. al* (Moccia *et al.*, 2017). As manchas (100x100 pixels) foram extraídas manualmente de 33 imagens laringoscópicas de banda estreita de 33 diferentes pacientes afetados por CECL, diagnosticados após exame histopatológico.

O banco de imagens original está particionado em três pastas, cada uma contendo 440 imagens. Cada uma das pastas está dividida em quatro classes, cada uma composta de 110 imagens focalizadas nas regiões de interesse: TS (tecido saudável), TVH (tecido com vasos hipertróficos), TLE (tecido com leucoplasia) e TACI (tecido com alças capilares intrapapilares), conforme amostra representada na Figura 3.2.



**Figura 3.2.** Amostra do dataset de imagens do tecido da laringe: (a) tecido com vasos hipertróficos, (b) tecido saudável, (c) tecido com alças capilares intrapapilares, (d) tecido com leucoplasia.

As imagens pertencentes ao *dataset* foram replicadas e redimensionadas para o tamanho de 28 por 28 pixels. Por fim, foram redistribuídas em três subconjuntos: treinamento, validação e teste. Na Tabela 3.1 é demonstrada divisão adotada para a realização da clas-

sificação com cada um dos tipos de classificadores.

**Tabela 3.1.** Organização do Dataset sem Aumento de Dados

Classificador	Treinamento	Validação	Teste
CNN (VGG 16, VGG19 e Inception V3)	1.188(90%)	0	132(10%)
CapsNet	1.056(80%)	132(10%)	132(10%)

Para a implementação do classificador CNN sem a adoção de *data augmentation*, optou-se pela segregação do conjunto de imagens em um grupo de treinamento, contendo 90% do total de imagens disponíveis, e os 10% restantes para teste.

A segregação do conjunto de imagens para implementação do classificador CapsNet antes do aumento de dados observou a proporção de 80% para treinamento, 10% para validação e 10% para teste, proporção comumente encontrada na literatura correlata à redes neurais artificiais. A distribuição do quantitativo de dados que serão utilizados nas etapas de treinamento, validação e teste é uma definição importante para o desempenho de algoritmos de aprendizagem de máquina.

### 3.2 *Data Aumentation* COM REDE GENERATIVA DCGAN

Após a obtenção das imagens foi adotado um recurso de *data augmentation* para aumentar a quantidade de imagens disponíveis para treinamento dos classificadores baseados em aprendizado de máquina: a partir do dataset original, foi desenvolvido um segundo banco de imagens ampliado, contendo, além das imagens originais, imagens artificiais geradas por uma GAN, implementada em linguagem Python na plataforma Colab.

A GAN foi construída utilizando o modelo de rede generativa adversarial convolucional profunda, *Deep Convolutional Generative Adversarial Network*, composta por uma rede geradora e outra discriminadora. As arquiteturas adotadas para a rede geradora e da rede discriminadora constam descritas detalhadamente nas tabelas 3.2 e 3.3.

Para geração das imagens artificiais, foram utilizadas 440 imagens pertencentes ao dataset original, observada a seguinte distribuição: 110 imagens de tecido saudável, 110 contendo leucoplasia, 110 referentes a vasos hipertróficos e 110 contendo alças intrapapilares.

As imagens geradas sinteticamente foram capturadas após o treinamento da rede DCGAN. O treinamento foi realizado por quantidade variável de épocas para cada categoria, até que o loss médio da rede discriminatória apresentasse valores médios das últimas 6 épocas próximos ao valor de 0.2 e a verificação empírica das amostras obtidas indicasse alto grau de semelhança com as amostras originais, conforme gráficos exibidos na figura

**Tabela 3.2.** Arquitetura da rede geradora DCGAN

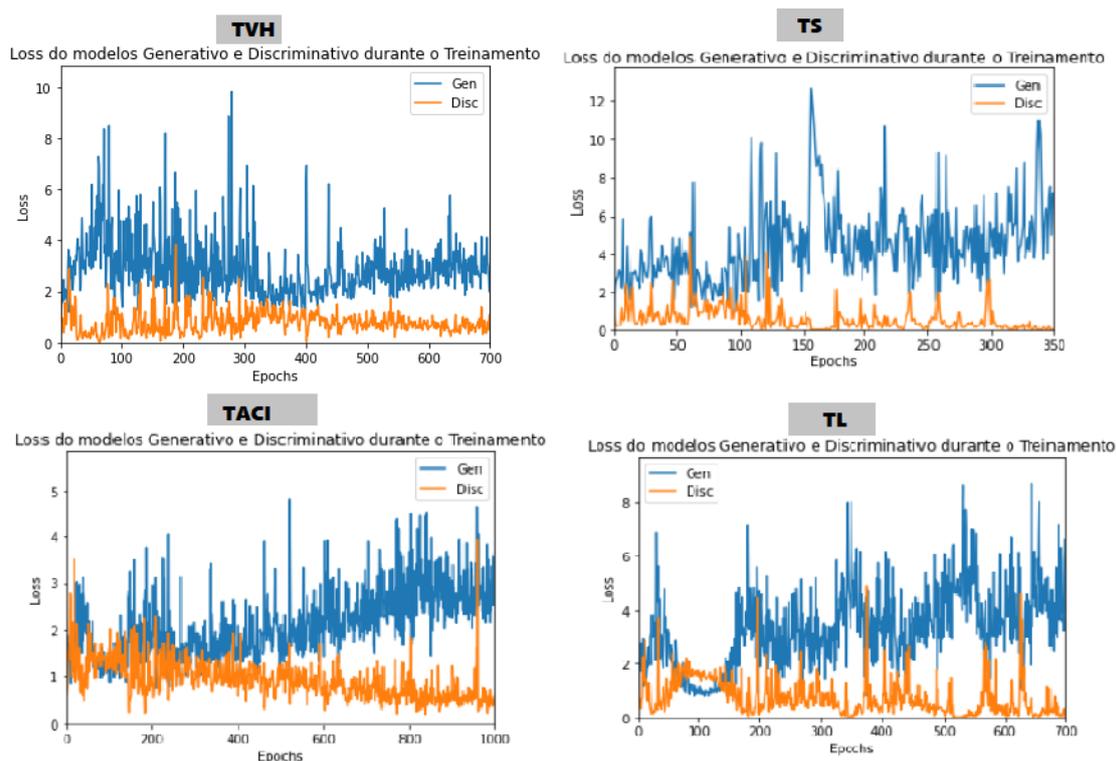
Camada	Formato de Saída	Parâmetros
Dense	(None, 4096)	4*4*256, activation = relu
Reshape	(None, 4, 4, 256)	formato = (4,4,256)
UpSampling 2D	(None, 8, 8, 256)	Redimensiona o formato para 8x8
Conv 2D	(None, 8, 8, 256)	kernel size = 3
BatchNormalization	(None, 8, 8, 256)	momentum = 0,2
Activation	(None, 8, 8, 256)	relu
UpSampling 2D	(None, 16, 16, 256)	Redimensiona o formato para 16x16
Conv2D	(None, 16, 16, 256)	kernel size = 3
BatchNormalization	(None, 16, 16, 256)	momentum = 0,2
Activation	(None, 16, 16, 256)	relu
UpSampling 2D	(None, 32, 32, 256)	Redimensiona o formato para 32x32
Conv2D	(None, 32, 32, 128)	kernel size = 3
BatchNormalization	(None, 32, 32, 128)	momentum = 0,2
Activation	(None, 32, 32, 128)	relu
UpSampling 2D	(None, 64, 64, 128)	Redimensiona o formato para 64x64
Conv2D	(None, 64, 64, 128)	kernel size = 3
BatchNormalization	(None, 64, 64, 128)	momentum = 0,2
Activation	(None, 64, 64, 128)	relu
Conv2D	(None, 64, 64, 3)	kernel size = 3
Activation	(None, 64, 64, 3)	tanh

**Tabela 3.3.** Arquitetura da rede discriminativa DCGAN

Camada	Formato de Saída	Parâmetros
Conv2D	(None, 32, 32, 32)	units = 32, kernel size = 3 e strides = 2
LeakyReLU	(None, 32, 32, 32)	alpha = 0,2
Dropout	(None, 32, 32, 32)	dropout = 0,25
Conv2D	(None, 16, 16, 64)	units = 64, kernel size = 3 e strides = 2
ZeroPadding2D	(None, 17, 17, 64)	
BatchNormalization	(None, 17, 17, 64)	momentum = 0,7
LeakyReLU	(None, 17, 17, 64)	alpha = 0,2
Dropout	(None, 17, 17, 64)	dropout = 0,25
Conv2D	(None, 9, 9, 128)	units = 128, kernel size = 3 e strides = 1
BatchNormalization	(None, 9, 9, 128)	momentum = 0,7
LeakyReLU	(None, 9, 9, 128)	alpha = 0,2
Dropout	(None, 9, 9, 128)	dropout = 0,25
Conv2D	(None, 9, 9, 256)	units = 256, kernel size = 3 e strides = 1
BatchNormalization	(None, 9, 9, 256)	momentum = 0,7
LeakyReLU	(None, 9, 9, 256)	alpha = 0,2
Dropout	(None, 9, 9, 256)	dropout = 0,25
Conv2D	(None, 9, 9, 512)	units = 512, kernel size = 3 e strides = 1
BatchNormalization	(None, 9, 9, 512)	momentum = 0,7
LeakyReLU	(None, 9, 9, 512)	alpha = 0,2
Dropout	(None, 9, 9, 512)	dropout = 0,25
Flatten	(None, 41472)	—
Dense	(None, 1)	activation = sigmoid

3.3 e tabela 1.2.

A Figura 3.3 a seguir apresenta as curvas de perda (*loss*) obtidas pela rede geradora, em azul, e pela rede discriminadora, em cor laranja, da estrutura Rede Generativa Adversária Convolutiva Profunda (do inglês *Deep Convolutional Generative Adversarial Network*) (DCGAN) em cada época de treinamento, para cada uma das quatro categorias de imagens: tecido saudável (TS), tecido contendo leucoplasia (TLE), lesões em padrão alças capilares intrapapilares (TACI) e vasos hipertróficos (TVH).



**Figura 3.3.** Resultados da curva de perda (*loss*) da rede generativa e da rede discriminatória, a época, pelo modelo DCGAN.

Em que pese a arquitetura DCGAN apresentar uma maior estabilidade durante o treinamento se comparada à arquitetura GAN tradicional, ainda pode apresentar instabilidade (Radford *et al.*, 2016), especialmente quando treinada por longos períodos ou com escassa quantidade de dados de entrada. É possível notar, na Figura 3.3, que as curvas de perda da rede generativa e da rede discriminatória apresentam uma tendência de divergência no curso das etapas de treinamento da rede DCGAN para geração das imagens sintéticas de tecidos da laringe. Dessa forma, a análise empírica das amostras geradas pela DCGAN foi a estratégia adotada para validar as imagens sintéticas e incorporá-las ao banco de imagens ampliado utilizado nas etapas seguintes.

### 3.3 MINERAÇÃO DOS DADOS E CLASSIFICAÇÃO DAS IMAGENS

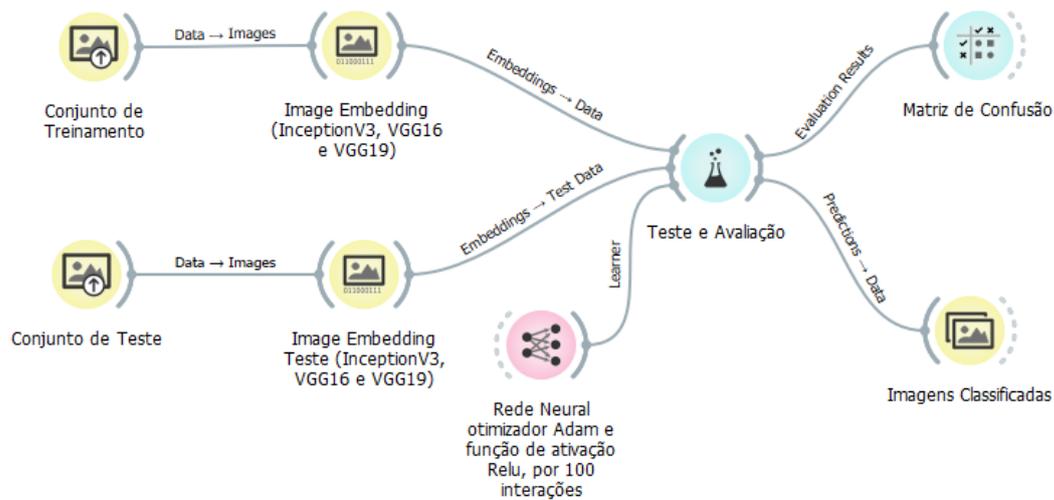
Para realizar a classificação das imagens, foi desenvolvido um classificador baseado em rede de cápsula, utilizado a plataforma Colab, e implementados três classificadores baseados em CNN com transferência de aprendizado, a saber VGG16, VGG19 e Inception V3, por meio dos componentes disponibilizados pelo complemento *Image Analytics* do software Orange Datamining.

Os classificadores baseados em modelos CNN foram implementados a partir da plataforma Orange Datamining, conforme *workflow* apresentado na Figura 3.4. Para realizar a classificação das imagens foram utilizados as seguintes ferramentas e parâmetros:

- *Import Images* - utilizado para importação das imagens pertencentes ao conjunto de treinamento e de teste;
- *Image Embedding* - incorporação de imagem através de rede neural profunda, permitindo a seleção dos parâmetros VGG16, VGG19 e InceptionV3;
- *Neural Network* - utilizado para implementação de um algoritmo perceptron multicamadas com *backpropagation* com os parâmetros: 100 neurônios por camada, o otimizador Adam, baseado em descida gradiente estocástica com regularização 0.0001, função ReLU, e treinamento por 100 iterações;
- *Test and Score* - realiza o treinamento conforme os parâmetros selecionados e exibe, dentre outras, as métricas de desempenho acurácia, precisão e f1-score;
- *Confusion Matrix* - exibição da quantidade de imagens classificadas, de forma correta e incorreta, pelo algoritmo em cada classe;
- *Image Viewer* - exibição das imagens e rótulos preditos a partir da classificação.

O classificador em rede de cápsula foi construído e implementado por meio da plataforma Google Colab, tendo por base o código disponibilizado com licença MIT, em @XifengGu CapsNet Keras (<https://github.com/XifengGuo/CapsNet-Keras>). O modelo apresentado originalmente utiliza 1.152 cápsulas primárias para imagens de entrada em preto e branco, com dimensões de 28 por 28 pixels, conforme demonstrado na Figura 3.5.

O modelo CapsNet foi adaptado para realizar a classificação a partir de imagens de entrada em cores rgb e também foi ajustada a taxa de aprendizado com otimizador Adam, para 0,0001. O algoritmo foi treinado primeiramente com o conjunto de imagens originais redimensionadas, e em seguida, foi treinado com o banco de imagens acrescido das 800 imagens geradas pela GAN, contendo ao todo 1.856 imagens no conjunto de treinamento.



**Figura 3.4.** Workflow utilizado para implementação dos Classificadores CNN no Software Orange Datamining.

As imagens geradas por GAN, com tamanho de saída 64 por 64 pixels, foram adicionadas ao conjunto de treinamento e redimensionadas para 28 por 28 pixels previamente ao treinamento.

O algoritmo CapsNet sem aumento de dados foi treinado por 400 épocas e o modelo CapsNet com aumento de dados foi treinado por 700 épocas, visto que a partir desse momento foi notada a ocorrência de overfitting, sobreajuste que ocorre quando um modelo de aprendizado de máquina apresenta bom desempenho no conjunto de treinamento, mas desempenho ruim nos dados de teste não vistos no treinamento.

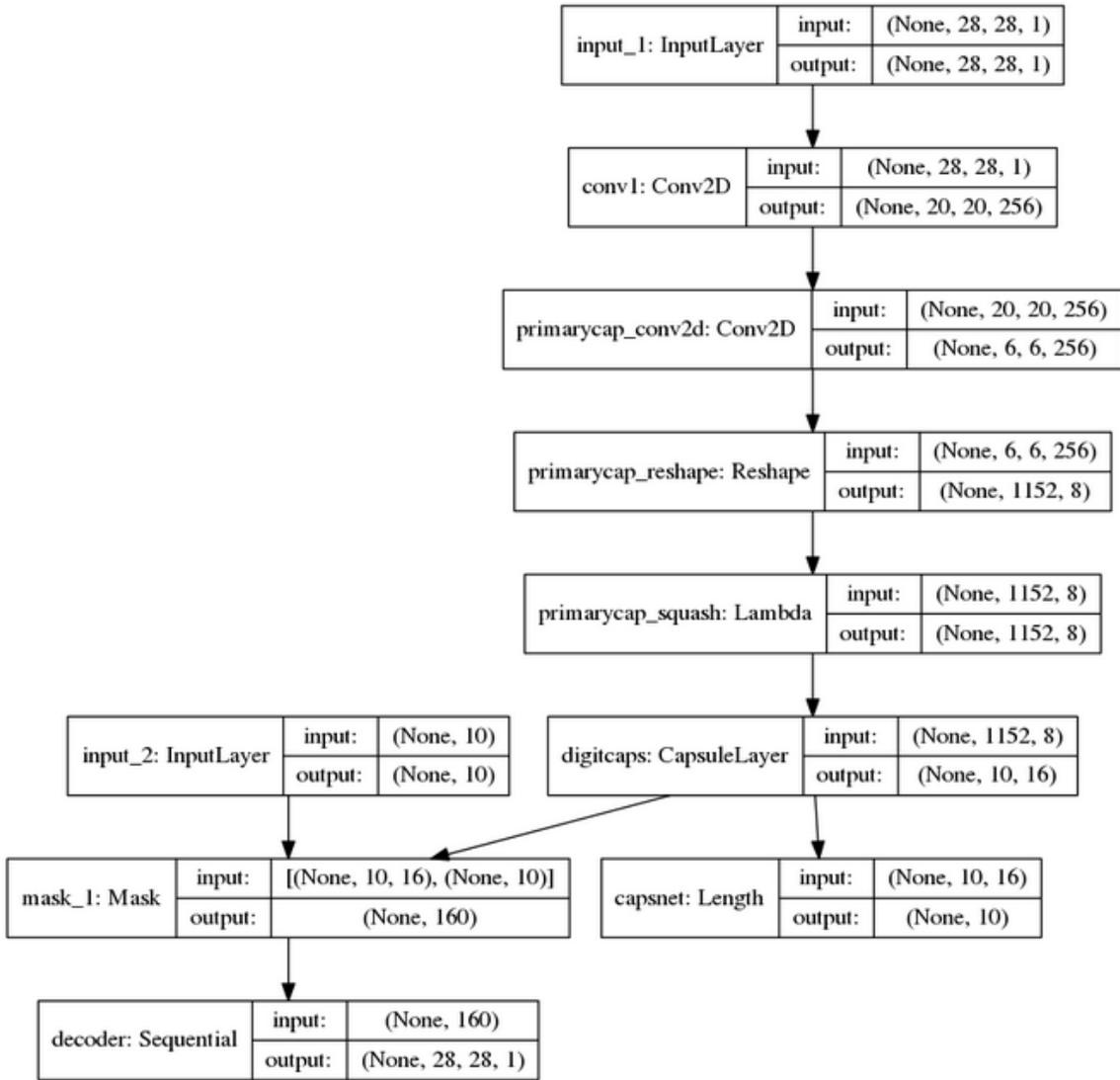
### 3.4 MÉTRICAS PARA AVALIAÇÃO DOS RESULTADOS

Os resultados obtidos por cada um dos classificadores implementados, CapsNet e CNN, foram analisados a partir dos seguintes métodos: matriz de confusão, métricas precisão, sensibilidade e F1-Score.

Uma matriz de confusão demonstra, em números absolutos ou percentuais, os acertos de um classificador em determinada classe.

A matriz de confusão dispõe as classes em linhas e colunas, e o cruzamento entre a linha e a coluna correspondentes a uma mesma classe representa a proporção de classificações corretas.

A Figura 3.6 apresenta um exemplo de uma matriz de confusão composta de três classes, A, B e C, cada uma delas composta de 20 imagens. As colunas destacadas em verde, na diagonal da matriz, representam a quantidade, em números absolutos, de



**Figura 3.5.** Arquitetura do modelo XifengGuo CapsNet Keras para imagens de entrada com dimensão 28x28 pixels.

imagens classificadas na categoria correta, e as destacadas em vermelho, as que foram preditas em classes divergentes.

A precisão está relacionada à proporção de amostras positivas classificadas corretamente entre todas as classificações positivas, enquanto a métrica de sensibilidade diz respeito ao percentual de acertos dentro da classe positiva.

A medida F1-Score é obtida a partir da média harmônica ponderada da precisão e da sensibilidade.

As equações para obter as métricas sensibilidade, precisão e F1-Score, constam especificadas respectivamente nas Equações 3.1, 3.2 e 3.3, onde VP são verdadeiros positivos, FP corresponde aos resultados falsos positivos, VN são verdadeiros negativos e FN são falsos negativos.

MATRIZ DE CONFUSÃO	CLASSE A	CLASSE B	CLASSE C	TOTAL
CLASSE A	10	5	5	20
CLASSE B	4	15	1	20
CLASSE C	6	2	12	20

Figura 3.6. Exemplo de uma matriz de confusão de três classes.

$$Sensibilidade = \frac{VP}{VP + FN} \quad (3.1)$$

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

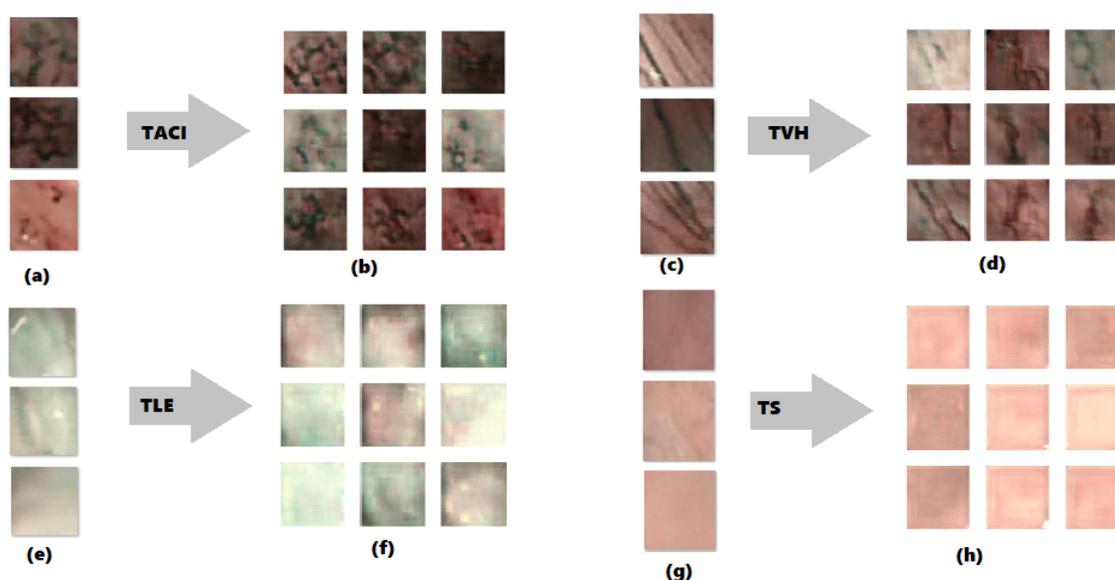
$$F1 - Score = \frac{2 * Precisão * Sensibilidade}{Precisão + Sensibilidade} \quad (3.3)$$

## 4 RESULTADOS E DISCUSSÃO

### 4.1 RESULTADO DA GERAÇÃO DAS IMAGENS

A aplicação da técnica de *data augmentation* utilizando uma rede neural do tipo Deep Convolutional GAN (DCGAN) obteve como resultado a geração de 200 imagens artificiais adicionais em cada uma das quatro classes do dataset original, conforme demonstrado na Tabela 4.1.

A Figura 4.1 apresenta uma amostra das imagens sintéticas obtidas a partir da adoção do modelo generativo.



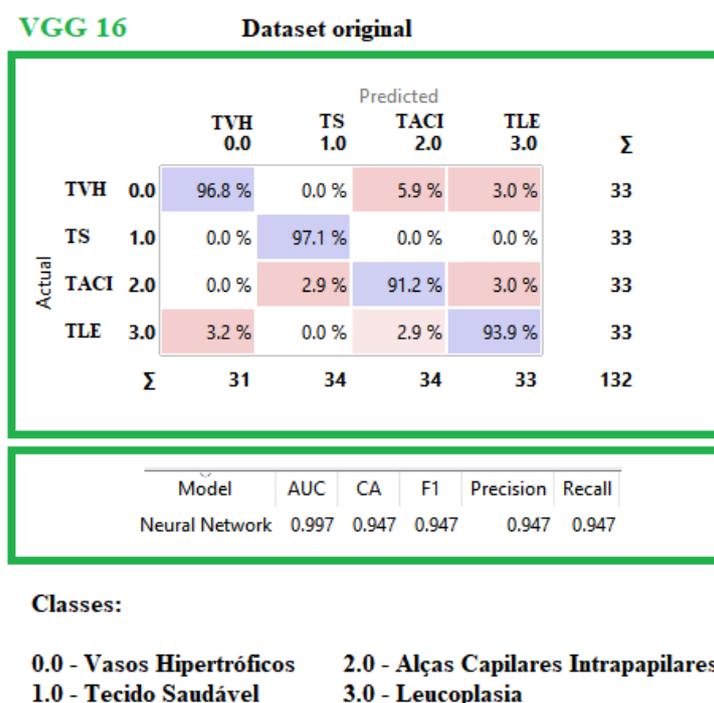
**Figura 4.1.** Amostra das imagens obtidas a partir da aplicação da técnica de data augmentation, onde (a), (c), (e) e (g) são amostras das imagens do dataset original e (b),(d),(f) e (f) correspondem as imagens geradas artificialmente com DCGAN.

**Tabela 4.1.** Resultado da Ampliação do Banco de Imagens com DCGAN

	Imagens Geradas	Epócas
TVH	200	700
TACI	200	1000
TLE	200	700
TS	200	350

## 4.2 RESULTADOS OBTIDOS PELO CLASSIFICADOR CNN ARQUITETURA VGG-16

O classificador CNN VGG16, após treinamento com as 1.188 imagens do dataset sem aumento de dados, obteve as três métricas avaliadas, precisão, sensibilidade e F1-Score correspondentes a 94,7% conforme demonstrado na Figura 4.2.



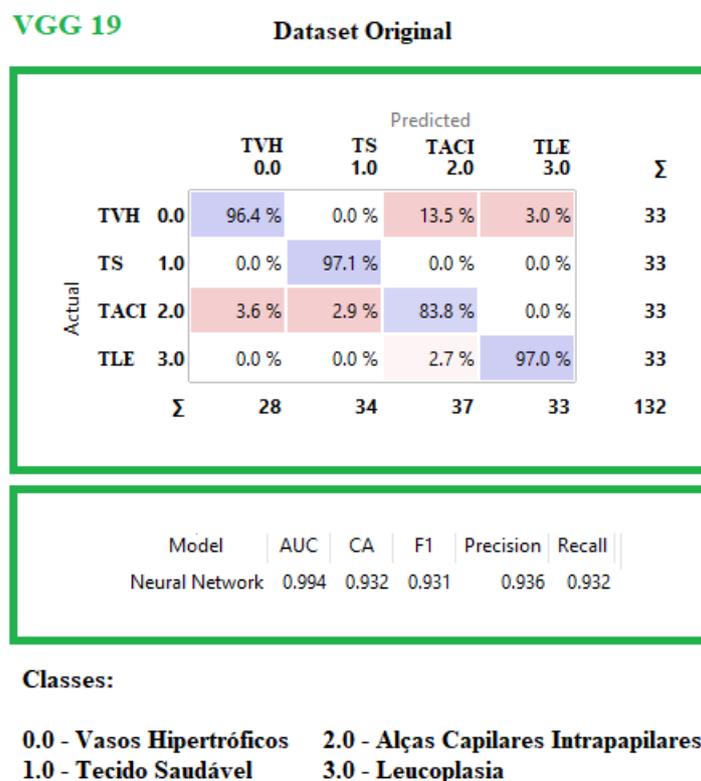
**Figura 4.2.** Resultados da matriz de confusão referente ao classificador CNN VGG 16.

A arquitetura CNN VGG16 obteve melhor desempenho na categorização das imagens contendo tecido saudável, com percentual de acerto nas predições de 97,1%. A presença de vasos hipertróficos foi identificada com percentual de acerto correspondente a 96,8%, tecido com lesões em padrão de alças capilares intrapapilares com 91,2% e leucoplasia

com 93,9%.

### 4.3 RESULTADOS OBTIDOS PELO CLASSIFICADOR CNN ARQUITETURA VGG-19

A rede VGG19 implementada no software ORANGE foi capaz de classificar corretamente 97,1% das imagens tecido saudável, 97% as imagens contendo leucoplasia, 96,4% as imagens de tecido com a presença de vasos hipertróficos e 83,8% as imagens contendo alças capilares intrapapilares.



**Figura 4.3.** Resultados da matriz de confusão referente ao classificador CNN VGG 19.

A CNN com transferência de aprendizado correspondente ao modelo VGG19 obteve precisão geral média de 93,6%. A sensibilidade média geral obtida pela arquitetura foi de 93,2% e o F1-Score de 93,1%.

#### 4.4 RESULTADOS OBTIDOS PELO CLASSIFICADOR CNN ARQUITETURA INCEPTION V3

O modelo baseado em Inception V3 conseguiu prever todas as 33 imagens pertencentes à classe tecido saudável de forma correta, conforme detalhado na matriz de confusão constante na Tabela 4.4. Dentre as imagens pertencentes à categoria leucoplasia, 97% foram classificadas corretamente. As imagens referentes às lesões em padrão de alças capilares intrapapilares e contendo vasos hipertróficos foram classificadas com aproveitamento de 94,1% e 96,9%, respectivamente.

**INCEPTION V3 Dataset Original**

		Predicted				$\Sigma$
		TVH 0.0	TS 1.0	TACI 2.0	TLE 3.0	
Actual	TVH 0.0	96.9 %	0.0 %	5.9 %	0.0 %	33
	TS 1.0	0.0 %	100.0 %	0.0 %	0.0 %	33
	TACI 2.0	0.0 %	0.0 %	94.1 %	3.0 %	33
	TLE 3.0	3.1 %	0.0 %	0.0 %	97.0 %	33
$\Sigma$		32	33	34	33	132

Model	AUC	CA	F1	Precision	Recall
Neural Network	0.999	0.970	0.970	0.970	0.970

**Classes:**

**0.0 - Vasos Hipertróficos      2.0 - Alças Capilares Intrapapilares**  
**1.0 - Tecido Saudável        3.0 - Leucoplasia**

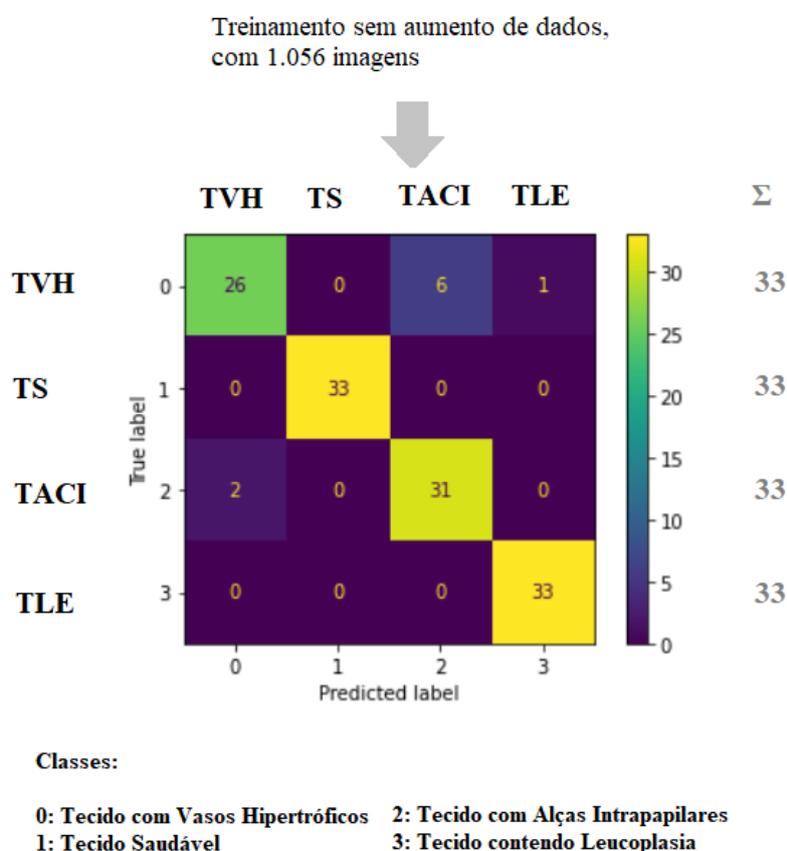
**Figura 4.4.** Resultados da matriz de confusão referente ao classificador CNN Inception V3.

A precisão global média obtida pelo modelo VGG 19 considerando todas as classes foi de 97%, assim como a sensibilidade média e a métrica F1-Score.

## 4.5 RESULTADOS OBTIDOS COM CAPSNET

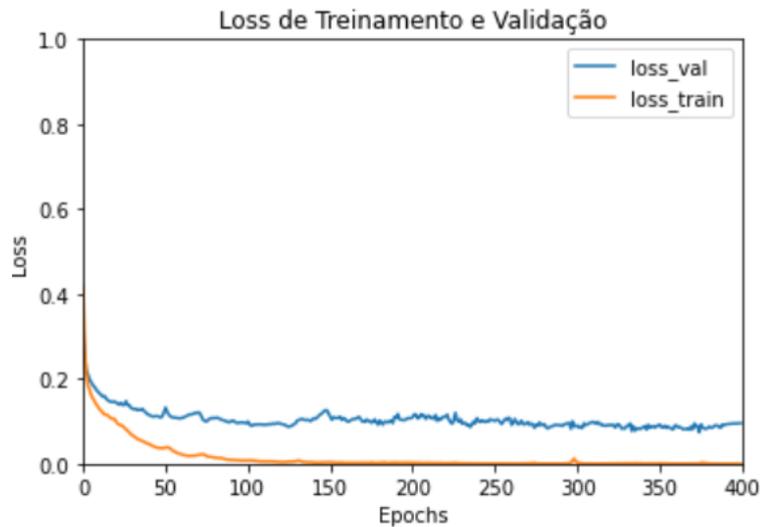
O classificador baseado em rede de cápsula treinado com o banco de imagens sem aumento de dados, composto por um conjunto de imagens de 1.056 utilizadas para treinamento e 132 imagens para validação, apresentou precisão de média de 93,4% na predição das imagens componentes do conjunto de teste, sensibilidade de 93,2% e estatística f1-score médio no valor de 93,1%.

A Figura 4.5, a seguir, apresenta a matriz de confusão contendo a quantidade de imagens classificadas em cada categoria.



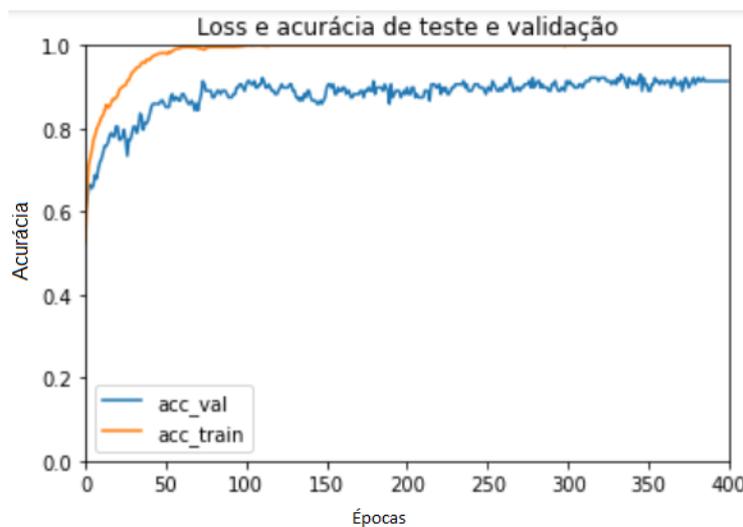
**Figura 4.5.** Resultados da matriz de confusão referente ao classificador baseado em rede de cápsulas.

O treinamento da rede foi realizado até 400 épocas, momento em que a curva de perda estabilizou. Na figura 4.6 a seguir, é demonstrada a evolução da curva de perda de treino e validação para o classificador CapsNet durante as 400 épocas do treinamento sem a adição de imagens artificiais geradas por GAN.



**Figura 4.6.** Evolução do Loss de Treinamento e Validação por época para a classificação CapsNet.

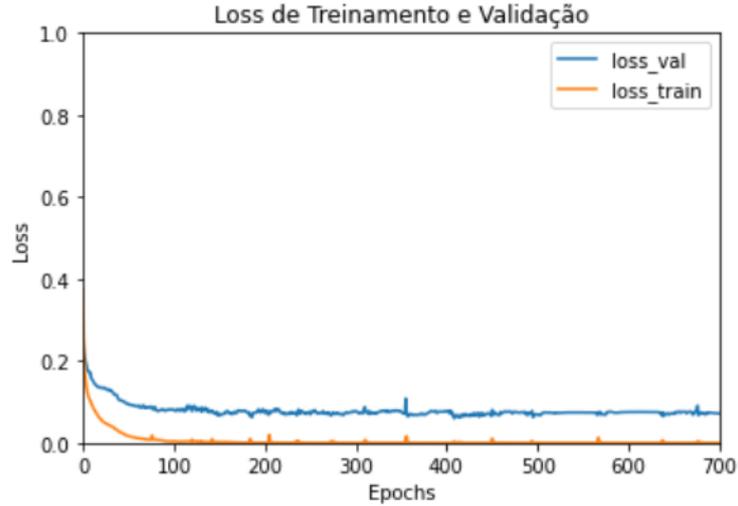
A Figura 4.7 exprime a acurácia obtida pelo classificador durante o treinamento e validação obtidos em cada época.



**Figura 4.7.** Evolução da Acurácia de Treino e Validação por época para a classificação com CapsNet.

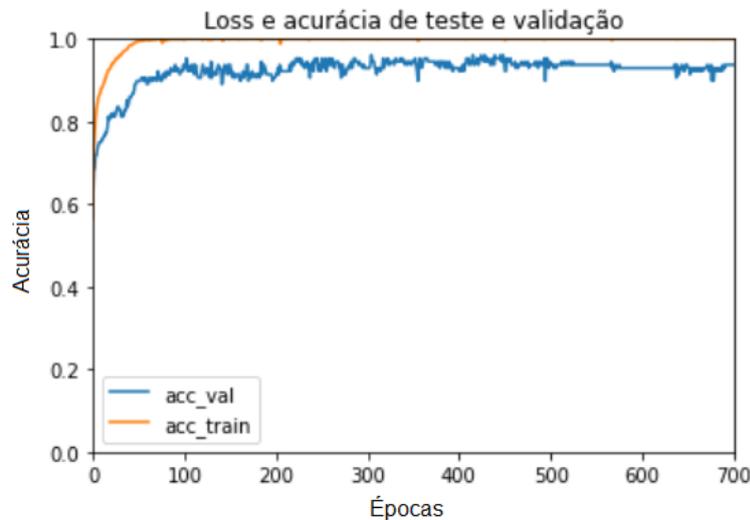
Após o treinamento utilizando o banco de imagens aumentado com imagens sintéticas geradas por DCGAN, no total de 1.856 imagens utilizadas para treino e 132 imagens para validação, o desempenho da rede CapsNet obteve melhoria das três estatísticas avaliadas, com precisão, sensibilidade e f1-score médios, respectivamente, correspondentes aos valores 94,72%, 94,7% e 94,7%.

A Figura 4.8 demonstra a evolução da curva de perda durante as etapas de treinamento e validação obtidos ao longo de 700 épocas pelo classificador CapsNet utilizando o conjunto de treinamento ampliado com *data augmentation* por GAN.



**Figura 4.8.** Evolução do Loss de Treinamento e Validação por época para a classificação CapsNet com data augmentation por GAN.

A Figura 4.9 apresenta a acurácia obtida pelo modelo CapsNet com o dataset aumentado com GAN.



**Figura 4.9.** Loss e Acurácia de teste por época para a classificação com CapsNet com data augmentation por GAN.

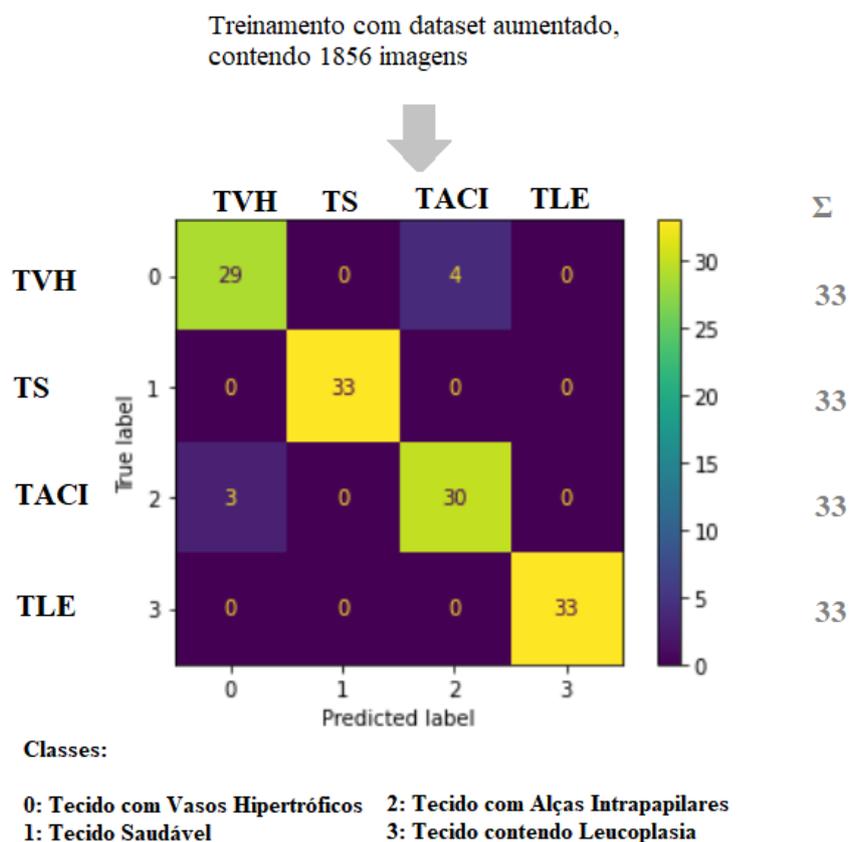
É relevante notar, no entanto, que ao comparar as curvas de acurácia obtidas pelas redes CapsNet dos gráficos constantes nas Figuras 4.7 e 4.9 é possível observar um descolamento entre os valores alcançados pelo modelo durante a etapa de treinamento e validação.

Essa divergência entre o desempenho obtido, que alcança 100% de acurácia durante o treinamento mas que decai de forma significativa durante a validação, sinaliza a presença de um fenômeno conhecido por *overfitting* ou sobreajuste, em que o modelo consegue

assimilar bem os dados de treinamento mas encontra dificuldade para generalizar os dados novos, ainda não vistos.

O sobreajuste pode ocorrer em situações que a amostra dos dados de entrada é muito pequena ou em que os dados apresentam pouca variabilidade. Mesmo adotadas medidas para ampliar o banco de imagens utilizadas para treinamento, ainda não foi possível sanar completamente a ocorrência de sobreajuste no modelo CapsNet.

Apesar do sobreajuste presente no modelo, adoção do recurso de aumento de dados com DCGAN combinada ao classificador baseado em cápsula aumentou, o desempenho nas três estatísticas avaliadas. A métrica F1-score registrou acréscimo de 1,6 pontos percentuais. A sensibilidade obtida com a técnica de aumento de dados foi 1,5 pontos percentuais maior que a classificação utilizando o conjunto de treinamento original, enquanto a precisão teve acréscimo de 1,32 pontos percentuais.



**Figura 4.10.** Resultados da matriz de confusão referente ao classificador baseado em rede de cápsulas, com aumento de dados.

Os valores de sensibilidade, precisão e F1-Score obtidos pelo modelo CapsNet na rotulação de cada classe quando treinado com a base de imagens original e com o base ampliada constam descritos, respectivamente, nas tabelas 4.3, 4.4 e 4.5.

Conforme Tabela 4.3, as classes tecido saudável e leucoplasia foram classificadas com sensibilidade correspondente a 100% tanto com a utilização do banco de imagens original como com o aumento de dados. Para a classe vasos hipertróficos, a sensibilidade obtida pela CapsNet foi maior com a utilização das imagens geradas artificialmente, correspondendo a 78,8% sem aumento e 87,9% com o aumento de dados, 10,1% superior. Somente para a classe alças intrapapilares a sensibilidade obtida foi inferior com *data augmentation*, com redução de 93,9% para 90,9%.

**Tabela 4.2.** Sensibilidade na classificação com CapsNet, por classe

Classe	Dataset Original	Dataset Aumentado com GAN
Vasos Hipertróficos	0.788	0.879
Tecido Saudável	1.000	1.000
Alças Capilares Intrapapilares	0.939	0.909
Leucoplasia	1.000	1.000

A precisão obtida pelo modelo foi adoção de data augmentation com DCGAN foi igual ou superior para todas as classes. A predição das imagens contendo leucoplasia alcançou 100% de precisão, 2,9% pontos percentuais a mais do que a classificação sem o aumento de dados. As imagens que continham tecido saudável foram classificadas com precisão de 100%. Para a classificação das demais lesões, vasos hipertróficos e alças capilares intrapapilares, o classificador CapsNet alcançou 90,6% e 88,2% de precisão respectivamente, no melhor desempenho, com o acréscimo das imagens sintéticas.

**Tabela 4.3.** Precisão na classificação com CapsNet, por classe

Classe	Dataset Original	Dataset Aumentado com GAN
Vasos Hipertróficos	0.928	0.906
Tecido Saudável	1.000	1.000
Alças Capilares Intrapapilares	0.838	0.882
Leucoplasia	0.971	1.000

A métrica F1-Score obtida também foi melhor na categorização de todas as classes com a utilização do conjunto de imagens ampliado por DCGAN. Quando utilizado o conjunto de dados aumentado no treinamento, CapsNet alcançou F1-Score de 100% na detecção de leucoplasia, 1,5 pontos percentuais a mais do que com o conjunto original, 100% na classificação de imagens do tecido saudável, 89,2% na classificação dos vasos hipertróficos, 89,5% para alças capilares intrapapilares.

**Tabela 4.4.** F1-Score na classificação com CapsNet, por classe

Classe	Dataset Original	Dataset Aumentado com GAN
Vasos Hipertróficos	0.852	0.892
Tecido Saudável	1.000	1.000
Alças Capilares Intrapapilares	0.886	0.895
Leucoplasia	0.985	1.000

## 4.6 CONSIDERAÇÕES FINAIS

Na tabela 4.5 constam os resultados obtidos por cada um dos classificadores avaliados.

**Tabela 4.5.** Resumo dos Resultados

Arquitetura	F1-score	Precisão	Sensibilidade	Imagens Treinamento
CNN VGG 16	0.947	0.947	0.947	1.188
CNN VGG 19	0.931	0.936	0.932	1.188
CNN Inception v3	0.970	0.970	0.970	1.188
CapsNet	0.931	0.934	0.932	1.056 + 132 validação
CapsNet + DCGAN	0.947	0.947	0.947	1.856 + 132 validação

De todos os modelos implementados, CNN Inception V3 aplicada ao conjunto de dados original foi o método que obteve melhor desempenho nas três métricas avaliadas, alcançando valores superiores a 97%.

O classificador CapsNet combinado ao recurso de *data augmentation* com imagens artificiais geradas por GAN apresentou o segundo melhor desempenho, com 94,7% de sensibilidade, f1-score e precisão geral, junto ao modelo CNN VGG16, que alcançou os mesmos valores.

Os resultados obtidos pelo modelo CNN VGG19 e CapsNet sem aumento de dados alcançaram valores médios próximos a 93%.

Em comparação com os estudos relativos à classificação multiclasse de lesões da laringe em imagens obtidas por NBI, os resultados obtidos pelo método Inception V3 foram os que mais se aproximaram do estado da arte (98%). O novo método CapsNet com aumento de dados por GAN também obteve resultados satisfatórios, próximos a 95%. O Quadro 4.1 demonstra a seguir a disposição dos melhores resultados obtidos para a métrica sensibilidade considerando o presente estudo e os resultados encontrados nos trabalhos relacionados.

**Quadro 4.1.** Análise dos resultados junto aos trabalhos relacionados

Autor(es)	Tipo de Classificação	Resumo	Sensibilidade média
(Araújo <i>et al.</i> , 2019)	Multiclasses	SVM e CNN	98%
(Moccia <i>et al.</i> , 2017)	Multiclasses	SVM	98%
<b>Costa, 2023*</b>	<b>Multiclasses</b>	<b>CNN Inception V3 utilizando programação visual com Orange Datamining</b>	<b>97%</b>
(Kwon <i>et al.</i> , 2022)	Binária	Árvores de Decisão	97%
(Inaba <i>et al.</i> , 2020)	Binária	Retina Net	95,5%
(Azam <i>et al.</i> , 2022b)	Multiclasses	SegMENT	95%
<b>Costa, 2023*</b>	<b>Multiclasses</b>	<b>CapsNet e DCGAN</b>	<b>94,7%</b>
(Xiong <i>et al.</i> , 2019)	Multiclasses	CNN	92,2%
(He <i>et al.</i> , 2021)	Binária	CNN	90,1%
(Esmacili <i>et al.</i> , 2020)	Binária	RF	84,6%
(Azam <i>et al.</i> , 2022a)	Binária	CNN YOLO	62%

Fonte: elaboração própria.

## 5 CONCLUSÃO

O câncer de laringe é uma patologia que pode comprometer severamente a qualidade de vida das pessoas, levando frequentemente à óbito quando identificado em estágio avançado. A presença de determinadas lesões na laringe pode sinalizar o desenvolvimento de tecido pré-cancerígeno, portanto, a identificação precoce desses sinais é importante para assegurar uma melhor eficácia do tratamento.

Este trabalho apresentou uma proposta de aplicação de redes neurais artificiais profundas para automatização da classificação de lesões na laringe a partir de imagens obtidas por laringoscopia com a técnica de imagens de banda estreita.

Foi proposto um novo método de classificação de imagens da laringe, que consistiu na utilização de um classificador baseado em rede de cápsulas aliado ao recurso de aumento de dados com DCGAN, apresentou desempenho satisfatório em classificar lesões da laringe em imagens obtidas por NBI, com sensibilidade médio geral de 94,7%. O modelo foi capaz de classificar as imagens contendo tecido saudável e leucoplasia com sensibilidade, precisão e f1-score médios de 100%. A presença de vasos hipertróficos foi classificadas com precisão de 90,6%, sensibilidade 87,9% e F1-Score de 89,2%. As lesões em padrão de alças capilares intrapapilares foram classificadas com precisão de 90,9%, sensibilidade 88,2% e F1-Score de 89,5%. No entanto, foi identificada a presença de sobreajuste do modelo.

A adoção do recurso de aumento de dados com acréscimo de imagens médicas sintéticas geradas pela DCGAN influenciou positivamente o desempenho do classificador baseado em rede de cápsula em todas as métricas avaliadas. A métrica F1-Score aumentou em 1,6 pontos percentuais com o recurso de *data augmentation*, enquanto a sensibilidade foi aprimorada em 1,5 pontos percentuais, e a precisão em 1,32. A técnica de *data augmentation* proposta, referente à geração de imagens sintéticas com o auxílio de uma DCGAN, foi capaz de melhorar o desempenho do classificador baseado em rede de cápsula em todas as medidas avaliadas na classificação de lesões da laringe, ainda que presente alguma instabilidade no modelo generativo utilizado.

Métodos baseados em RNA são importantes ferramentas para auxiliar o diagnóstico de doenças em exames de imagem. A nova solução desenvolvida, que combina redes de

cápsula e GAN, pode representar uma alternativa interessante para o processo de classificação automatizada de imagens médicas com redes neurais artificiais, especialmente nas situações que envolvem doenças raras, em que as imagens disponíveis para treinamento dos algoritmos de aprendizagem de máquina podem se apresentar escassas, entretanto, carece de aprimoramento para evitar a ocorrência sobreajuste e reduzir a instabilidade do modelo generativo.

Também foi proposta a implementação de classificadores baseados em CNN convencionais a partir da ferramenta de programação visual e mineração de dados Orange Data Mining. O modelo baseado em Inception V3, implementado com programação visual, obteve o melhor desempenho geral, superando inclusive o modelo CapsNet classificação das imagens com sensibilidade, precisão e F1-Score médios equivalentes a 97%, próximos ao estado da arte (98%).

Este estudo contribuiu para o avanço do conhecimento em tecnologias relacionadas à saúde ao explorar novos métodos para o auxílio do diagnóstico de lesões precursoras de câncer da laringe. A principal contribuição deste estudo, portanto, consistiu na apresentação de uma solução de inteligência artificial baseada na implementação de redes neurais artificiais convolucionais através da ferramenta de programação visual, que demonstrou-se uma técnica eficiente na classificação de imagens médicas, permitindo a classificação de lesões da laringe com 97% de precisão e sensibilidade, tendo por diferencial a baixa complexidade de utilização, baixo custo operacional e pouca dependência de conhecimentos avançados de programação do usuário.

## 6 LIMITAÇÕES E TRABALHOS FUTUROS

Este estudo se concentrou na aplicação da técnica de aumento da base de dados com GAN utilizando apenas um modelo baseado em rede de cápsula para classificação automatizada de lesões da laringe. Portanto, são necessários mais estudos para analisar o impacto desse método de aumento de dados quando combinado com outros modelos de classificadores.

Os classificadores foram treinados com imagens de baixa resolução, no tamanho de 28 pixels por 28 pixels. É relevante que sejam realizados mais estudos, com a adaptação do modelo de forma a possibilitar o treinamento da rede com imagens em resoluções maiores, que permitam que a RNA identifique mais detalhes.

Uma limitação identificada neste estudo consistiu na divisão do *dataset* utilizado pelos classificadores CNN implementados com programação visual, visto que não foram reservadas imagens para compor o subconjunto de validação. Assim, há uma oportunidade de melhoria dos classificadores CNN com a destinação de uma parte das imagens para realizar a validação do modelo.

O modelo DCGAN utilizado neste estudo apresentou instabilidade durante o treinamento, haja vista a reduzida amostra dos dados entrada, ainda assim, a técnica foi capaz de melhorar o desempenho do classificador CapsNet. Assim, o ajuste fino para aumentar a estabilidade do modelo DCGAN durante o treinamento, bem como a adição de mecanismo para automatizar a validação das amostras geradas pela rede DCGAN com o auxílio de métricas de avaliação da fidedignidade com os dados reais, tais como *Frechet Inception Distance* (FID), são sugestões para trabalhos futuros.

A presença de um descolamento visível entre as curvas de acurácia de validação e de treinamento originadas na classificação realizada com os modelos de redes de cápsulas revelou indícios de *overfitting*. Assim, a utilização de uma base de dados maior para o treinamento e a implementação de mecanismos de controle automatizado de sobreajuste nos modelos CapsNet são melhorias sugeridas para aprimorar o desempenho dos classificadores CapsNet implementados na plataforma Colab em trabalhos futuros.

Adicionalmente, são apresentadas, como sugestões de trabalhos futuros relacionados aos assuntos explorados nesta pesquisa, as seguintes propostas:

- robustecer o mecanismo de data augmentation com imagens geradas porGAN, com a adaptação da rede generativa para gerar imagens em maior resolução, e refazer a classificação utilizando imagens de entrada em maiores dimensões pode melhorar o desempenho obido pelo classificador baseado em rede de cápsula,vez que permitirão a captura de maior quantidade de detalhes presentes nas imagens das lesões do tecido da laringe;
- ampliar de forma mais ostensiva o conjunto de dados para treinamento, com a geração de um maior número de imagens sintéticas pela rede DCGAN e complementar com a utilização de outros recursos clássicos de *data augmentation* como rotação e redimensionamento;
- integrar o modelo de inteligência artificial proposto em uma interface gráfica de usuário, como um sistema web ou aplicativo móvel;
- aplicar combinação da técnica de aumento da base de dados com GAN combinada com classificador baseado em rede de cápsula para avaliar o desempenho desse modelo aprendizagem de máquina na tarefa de classificação de outros tipos de imagens médicas, referentes a outras patologias;
- aplicar a técnica de aumento de dados com GAN em outros modelos de aprendizagem de máquina para classificar imagens médicas relativos às lesões da laringe.

## LISTA DE REFERÊNCIAS

- Abadi, M., Barham, P., Chen, J., *et al.* (2016). Tensorflow: A system for large-scale machine learning.
- Abreu, G. K. N., de Abreu, D. D. C., Leite, A. L., e Moreira, J. A. (2020). Análise dos fatores influentes no surgimento de câncer de faringe na população mundial. *Research, Society and Development*, 9.
- Afshar, P., Oikonomo, A., Naderkhani, F., *et al.* (2020). 3d-mcn: A 3d multi-scale capsule network for lung nodule malignancy prediction. *Scientific Reports*, 10.
- Aloysius, N. e Geetha, M. (2018). A review on deep convolutional neural networks. volume 2018-January, páginas 588–592. Institute of Electrical and Electronics Engineers Inc.
- Amar, A. *et al.* (2010). Retardo diagnóstico e terapêutico em pacientes com câncer da laringe em hospital público de referência. *Brazilian Journal of Otorhinolaryngology*, 76.
- Araújo, T., Santos, C., Momi, E. D., e Moccia, S. (2019). Learned and handcrafted features for early-stage laryngeal scc diagnosis. *Medical and Biological Engineering and Computing*, 57.
- Azam, A. M. *et al.* (2022a). Deep learning applied to white light and narrow band imaging videolaryngoscopy: Toward real-time laryngeal cancer detection. *LARYNGOSCOPE*, 132(9):1798–1806.
- Azam, M. A. *et al.* (2022b). Videomics of the upper aero-digestive tract cancer: Deep learning applied to white light and narrow band imaging for automatic segmentation of endoscopic images. *FRONTIERS IN ONCOLOGY*, 12.
- Borges, R. R., Iarosz, K. C., Batista, A. M., Caldas, I. L., Borges, F. S., e Lameu, E. L. (2015). Sincronização de disparos em redes neuronais com plasticidade sináptica. *Revista Brasileira de Ensino de Física*, 37.
- Chen, M. *et al.* (2019). A morphological classification for vocal fold leukoplakia. *Brazilian Journal of Otorhinolaryngology*, 85.

- Chollet, F. (2015). Keras documentation.
- Demšar, J., Curk, T., Erjavec, A., *et al.* (2013). Orange: Data mining toolbox in python. *Journal of Machine Learning Research*, 14.
- Dogantemur, S. *et al.* (2020). Assessment of hpv 16, hpv 18, p16 expression in advanced stage laryngeal cancer patients and prognostic significance: Assessment of hpv 16, hpv 18, p16 expression in laryngeal cancer. *Brazilian Journal of Otorhinolaryngology*, 86.
- dos Anjos, B., Jatobá, A., e Oliveira, M. (2019). Generative adversarial networks: a renewal for data augmentation in lung nodule classification.
- Esmaeili, N. *et al.* (2020). Laryngeal lesion classification based on vascular patterns in contact endoscopy and narrow band imaging: Manual versus automatic approach. *Sensors (Switzerland)*, 20.
- Fayyad, U., Piatetsky-Shapiro, G., e Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17.
- Filho, O. M. e Neto, H. V. (1999). *Processamento Digital de Imagens*. Editora Brasport, Rio de Janeiro, Brazil.
- Frid-Adar, M., Diamant, I., Klang, E., *et al.* (2018). Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification. *Neurocomputing*, 321.
- Galbusera, F., Niemeyer, F., Seyfried, M., *et al.* (2018). Exploring the potential of generative adversarial networks for synthesizing radiological images of the spine to be used in in silico trials. *Frontiers in Bioengineering and Biotechnology*, 6.
- Gale, N., Gnepp, D. R., Poljak, M., Strojan, P., Cardesa, A., Helliwell, T., Šifrer, R., Volavšek, M., Sandison, A., e Zidar, N. (2016). Laryngeal squamous intraepithelial lesions: An updated review on etiology, classification, molecular changes, and treatment.
- Godec, P., Pančur, M., Ilenič, N., *et al.* (2019). Democratized image analytics by visual programming through integration of deep models and small-scale machine learning. *Nature Communications*, 10.
- Gonzalez, R. e Woods, R. (2008). *Digital image processing*. Prentice Hall, Upper Saddle River, N.J.
- Gonzalez, R., Woods, R., e Eddins, S. (2004). Digital image processing using matlab - gonzalez woods and eddins.pdf.

- Haffeman, L. G. (2014). *An analysis of deep neural networks for texture classification*. Dissertação de mestrado, Programa de Pós-Graduação em Informática, Universidade Federal do Paraná, Setor de Ciências Exatas.
- He, Y. *et al.* (2021). A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis. *ANNALS OF TRANSLATIONAL MEDICINE*, 9(24).
- Hermans, R. (2006). Staging of laryngeal and hypopharyngeal cancer: Value of imaging studies. *European Radiology*, 16.
- Hesamian, M. *et al.* (2019). Deep learning techniques for medical image segmentation: Achievements and challenges. *Journal of Digital Imaging*, 32.
- Hinton, G. e Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. *Science*, 313.
- Inaba, A. *et al.* (2020). Artificial intelligence system for detecting superficial laryngopharyngeal cancer with high efficiency of deep learning. *Head and Neck*, 42.
- INCA, I. (2020). Câncer de laringe. <https://www.inca.gov.br/tipos-de-cancer/cancer-de-laringe>.
- INCA, I. (2023). Câncer de laringe. <https://www.inca.gov.br/sites/ufu.sti.inca.local/files/media/document/estimativa-2023.pdf>.
- Klimza, H., Jackowska, J., Piazza, C., Banaszewski, J., e Wierzbicka, M. (2019). The role of intraoperative narrow-band imaging in transoral laser microsurgery for early and moderately advanced glottic cancer. *Brazilian Journal of Otorhinolaryngology*, 85.
- Kosgiker, G. *et al.* (2021). Segcaps: An efficient segcaps network-based skin lesion segmentation in dermoscopic images. *International Journal of Imaging Systems and Technology*, 31.
- Kwon, I., Wang, S., Shin, S., Cheon, Y., Lee, B., Lee, J., Lim, D., Jo, C., Cho, Y., e Shin, B. (2022). Diagnosis of early glottic cancer using laryngeal image and voice based on ensemble learning of convolutional neural network classifiers. *Journal of Voice*.
- LaLonde, R. *et al.* (2021). Capsules for biomedical image segmentation. *Medical Image Analysis*, 68.
- LeCun, Y., Kavukcuoglu, K., e Farabet, C. (2010). Convolutional networks and applications in vision.

- Liu, W. *et al.* (2017). A survey of deep neural network architectures and their applications. *Neurocomputing*, 234.
- Lobo, L. (2017). Inteligência artificial e medicina. *Revista Brasileira de Educação Médica*, 41.
- Lukes, P. *et al.* (2013). Narrow band imaging (nbi) — endoscopic method for detection of head and neck cancer.
- Maciel, C. T. V., Leite, I. C. G., e Soares, T. V. (2010). Câncer de laringe: um olhar sobre a qualidade de vida. *Revista Interdisciplinar de Estudos Experimentais - Animais e Humanos (RIEE)*, 2.
- Matsugu, M., Mori, K., Mitari, Y., e Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. volume 16.
- Menezes, A. M. B., Horta, B. L., Oliveira, A. L. B., Kaufmann, R. A. C., Duquia, R., Diniz, A., Motta, L. H., Centeno, M. S., Estanislau, G., e Gomes, L. (2002). Risco de câncer de pulmão, laringe e esôfago atribuível ao fumo. *Revista de Saúde Pública*, 36.
- Moccia, S., Momi, E. D., Guarnaschelli, M., Savazzi, M., e Laborai, A. (2017). Confident texture-based laryngeal tissue classification for early stage diagnosis support. *Journal of Medical Imaging*, 4.
- Nemetz, M., Pontes, P., Vieira, V., e Yazaki, R. (2005). Configuração das pregas vestibulares à fonação em adultos com e sem disfonia. *Revista Brasileira de Otorrinolaringologia*, 71.
- Ni, X. *et al.* (2011). Endoscopic diagnosis of laryngeal cancer and precancerous lesions by narrow band imaging. *Journal of Laryngology and Otology*, 125.
- Oliveira, A. T. (2019). *Classificação automática de arritmias cardíacas usando uma combinação de redes neurais*. Tese de doutorado, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas - Unicamp.
- Panwar, A., Lindau, R., e Wieland, A. (2013). Management of premalignant lesions of the larynx.
- Park, J. *et al.* (2021). Laryngeal leukoplakia: State of the art review.
- Piazza, C., Bon, F. D., *et al.* (2012). Narrow band imaging in endoscopic evaluation of the larynx.

- Piazza, C. *et al.* (2010). Role of narrow-band imaging and high-definition television in the surveillance of head and neck squamous cell cancer after chemo- and/or radiotherapy. *European Archives of Oto-Rhino-Laryngology*, 267.
- Radford, A., Metz, L., e Chintala, S. (2016). Unsupervised representation learning with deep convolutional gans. *International Conference on Learning Representations*.
- Roesler, V., Kronbauer, A., Neto, M., *et al.* (2018). *Desenvolvendo Modelos de Deep Learning para Aplicações Multimídia no Tensorflow*.
- Rumelhart, D., Hinton, G., e Williams, R. (1986). Learning representations by back-propagating errors. *Nature*, 323.
- Sabour, S., Frosst, N., e Hinton, G. (2017). Dynamic routing between capsules. volume 2017-December.
- Sandfort, V., Yan, K., Pickhardt, P., e Summers, R. (2019). Data augmentation using generative adversarial networks (cycleGAN) to improve generalizability in ct segmentation tasks. *Scientific Reports*, 9.
- Santiago, M. B. (2015). *Análise de Proteoma no Carcinoma de Células Escamosas de Laringe*. Tese de doutorado, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas - Unicamp.
- Santos, A., Aires, K., Veras, R., Uchôa, V., e Santos, L. (2020). Uma abordagem de classificação de imagens dermatoscópicas utilizando aprendizado profundo com redes neurais convolucionais.
- Scalassara, P., Dajer, M., *et al.* (2009). Relative entropy measures applied to healthy and pathological voice characterization. *Applied Mathematics and Computation*, 207.
- Soares, E. B. (2009). *Caracterização Histomorfométrica e Índice Proliferativo (Ki-67) das Displasias Acentuadas/Carcinomas in situ nas Pregas Vocais*. Dissertação de mestrado, Programa de Pós-Graduação em Patologia, Universidade Federal de Pernambuco - UFPE, Recife.
- Uddin, S., Khan, A., Hossain, M., e Moni, M. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19.
- Unger, J. *et al.* (2015). A noninvasive procedure for early-stage discrimination of malignant and precancerous vocal fold lesions based on laryngeal dynamics analysis. *Cancer Research*, 75.
- WHO, W. H. O. (2020). The impact of cancer. <http://www.who.int/>.

- Xiong, H., Lin, P., Yu, J., Ye, J., Xiao, L., Tao, Y., Jiang, Z., Lin, W., Liu, M., Xu, J., Hu, W., Lu, Y., Liu, H., Li, Y., Zheng, Y., e Yang, H. (2019). Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine*, 48:92–99.
- Yadav, S. e Jadhav, S. (2019). Deep convolutional neural network based medical image classification for disease diagnosis. *Journal of Big Data*, 6.
- Zhao, D., Zhu, D., *et al.* (2018). Synthetic medical images using f and bganfor improved lung nodules classification bymulti-scale vgg16. *Symmetry*, 10.

# APÊNDICES

## APÊNDICE A

Os códigos e *datasets* utilizados para implementação dos algoritmos de aprendizado de máquina estão disponíveis no repositório público do GitHub ([https://github.com/analuzamfc/dissertacao\\_PPGEB](https://github.com/analuzamfc/dissertacao_PPGEB)):

- `Codigo_GAN.ipynb` - Código utilizado para a implementação da rede GAN com a biblioteca Keras em Python, adaptada da implementação original de Jeff Heaton, disponível no repositório público Gitee ([https://gitee.com/shark803/t81\\_558\\_deep\\_learning](https://gitee.com/shark803/t81_558_deep_learning));
- `Codigo_CapsNet.ipynb` - Código de uma rede de cápsula (CapsNet) adaptada da implementação original de Xifeng Guo, disponível no repositório público GitHub (<https://github.com/XifengGuo/CapsNet-Keras/>);
- `Fluxoorange.ows` - Workflow para implementação dos classificadores CNN;
- `Capsnet_dataset.zip` - Dataset composto das pastas "train", "validation" e "test", contendo, respectivamente 1.056, 132 e 132 imagens;
- `Cnn_dataset.zip` - Dataset composto das pastas "train" e "test", contendo, respectivamente 1.188 e 132 imagens;
- `Imagens_geradas_dcgan.zip` - Pasta contendo 800 imagens geradas pela rede GAN, divididas igualmente em quatro subpastas referentes às classes.

## APÊNDICE B

Slides utilizados na composição da apresentação do trabalho acadêmico.

Programa de Pós-Graduação em ENGENHARIA BIOMÉDICA

UnB

# Identificação de Lesões da Laringe em Imagens de Banda Estreita Utilizando Redes Neurais Artificiais e Programação Visual

Ana Luiza Moraes Fernandes da Costa

Orientador: Professor Dr. Gerardo Antônio Idrobo Pizo (UnB)  
Examinador: Professor Dr. Cristiano Jacques Miosso (UnB)  
Examinadora: Professora Dra. Luciana Correia Alves (UNICAMP)  
Examinador Suplente: Professor Dr. Renato Coral Sampaio (UnB)

Linha de Pesquisa: Tecnologias em Engenharia Biomédica

1

Figura 1. Slide 1

Programa de Pós-Graduação em ENGENHARIA BIOMÉDICA

UnB

## Estrutura da Apresentação

- Introdução e Justificativa**
  - Contextualização e definição do problema;
  - Objetivos;
  - Justificativa.
- Referencial Teórico**
  - Lesões na Laringe;
  - Processamento Digitais de Imagens;
  - Mineração de Dados e Classificação de Imagens Médicas com Redes Neurais;
  - Trabalhos Relacionados;
  - Considerações Finais.
- Materiais e Método**
  - Aquisição do *dataset*;
  - Ampliação do *dataset* utilizando redes generativas adversárias;
  - Mineração dos Dados e Classificação das Imagens;
  - Avaliação dos Resultados.
- Resultados, Discussão e Conclusões**

4/10

Figura 2. Slide 2

## Introdução e Justificativa

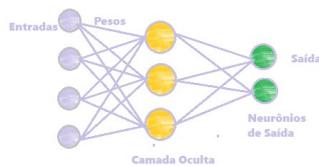
### Contextualização e problema de pesquisa

- Laringe → órgão musculocartilaginoso e multifuncional:
  - conduzir o ar durante a respiração;
  - proteger as vias aéreas inferiores;
  - função esfinteriana, evitando a entrada de água e alimentos; e inibindo a saída de ar dos pulmões (NEMETZ et al., 2005);
  - produção da voz humana.
- Lesões na laringe podem sinalizar o desenvolvimento de **Carcinoma Espinocelular da Laringe (CECL)** em estágio inicial.
- O câncer de laringe corresponde a cerca de **3% de todos os casos de neoplasia maligna** no mundo (DOGANTEMUR et. al., 2020), sendo que o CECL é a apresentação mais frequente: 98% dos casos de câncer de laringe.
- Estima-se mais de 184 mil ocorrências dessa neoplasia em 2020 (WORLD HEALTH ORGANIZATION, 2020) e **4.450 óbitos** decorrentes da doença no Brasil (INCA, 2020).



Figura 3. Slide 3

## Introdução e Justificativa



- Redes Neurais Artificiais:** baseadas na simulação computacional do funcionamento do cérebro humano, inspirando-se no processo de aprendizagem e transmissão de informações pelos neurônios biológicos.
- Redes Neurais Artificiais têm sido utilizadas para classificação de diferentes modalidades de imagens médicas.
- Rede de cápsula (Sabour, Frost e Hinton 2017)**
  - capaz de detectar objetos dispostos em uma imagem
  - conserva a informação sobre a relação dos objetos e da imagem como um todo
- Redes de Cápsula têm sido aplicadas no processamento de **imagens médicas**, obtendo **desempenho próximo ou superior** ao apurado por outros tipos de classificadores.



Figura 4. Slide 4

## Introdução e Justificativa

### Questões de Pesquisa

A identificação de lesões da laringe, que podem sinalizar CECL em estágio inicial, persiste um desafio para os especialistas clínicos.

- Um modelo de aprendizagem de máquina supervisionado baseado em arquitetura de **rede de cápsula (CapsNet)** é capaz de classificar carcinoma espinocelular da laringe em imagens obtidas por NBI, com desempenho próximo ou superior ao alcançado pelas principais arquiteturas de rede convolucional, exploradas na literatura científica?
- A adição de **imagens médicas sintéticas** geradas por **Rede Generativa Adversária (GAN)** ao **dataset** de treinamento influencia positivamente o desempenho do classificador baseado em rede de cápsula?



Figura 5. Slide 5

## Introdução e Justificativa

### Objetivos

#### Objetivo Geral:

O objetivo central deste estudo consistiu em desenvolver e avaliar o desempenho de um classificador baseado em **rede de cápsula** na tarefa de **classificação de lesões da laringe** em imagens de banda estreita obtidas por meio de exame de laringoscopia, utilizando recurso de aumento da base de dados com GAN.

#### Objetivos Específicos:

- Desenvolvimento de um classificador baseado em algoritmo de rede neural em cápsula;
- Adoção de técnica de *data augmentation* para ampliação da base de dados com imagens de tecido da laringe, geradas sinteticamente a partir de uma GAN (rede generativa adversária);
- Avaliação do desempenho do classificador baseado em cápsula, comparativamente a outros classificadores explorados na literatura relacionada;
- Avaliação do impacto do recurso de data augmentation com imagens geradas por GAN na performance dos classificador baseado em rede de cápsula;
- Desenvolvimento e avaliação de um classificador baseado em CNN utilizando software de programação visual.

6

Figura 6. Slide 6

## Introdução e Justificativa

- Diagnóstico **tardio** de CECL:
  - **tratamentos agressivos;**
  - **eleva as taxas de mortalidade;**
  - necessidade de **ressecção da laringe e das pregas vocais.**
- **Dificuldade** para o diagnóstico em estágio inicial:
  - lesões menores.
- Classificador baseado em **rede de cápsula**:
  - Auxiliar no diagnóstico de lesões em fase inicial
  - propiciar **intervenção precoce.**



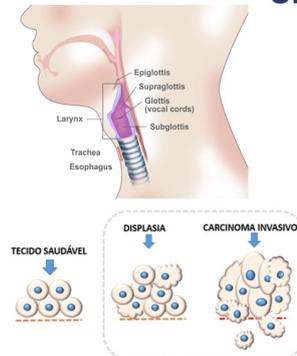
47

Figura 7. Slide 7

## Referencial Teórico

### Lesões da Laringe

- O Carcinoma Espinocelular da Laringe ocorre a partir do desenvolvimento anormal (displasia) das células escamosas que compõem a camada interna do revestimento da laringe.
- Principais Fatores de Risco:
  - *Tabagismo e alcoolismo*
- Um dos principais sinais indicativos de acometimento da laringe por carcinoma espinocelular consiste **na presença de lesões** como leucoplasia e vasos hipertróficos na mucosa.
- Sintomas como disfonia e disfagia aparecem em estágio avançado.
- **Imagem de banda estreita (NBI)** é uma técnica endoscópica moderna e utiliza filtros especiais que selecionam apenas os comprimentos de onda da luz verde e da luz azul, que permite melhor identificação de lesões pequenas na laringe.



48

Figura 8. Slide 8

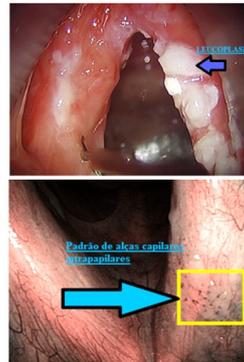
## Referencial Teórico

### Tecido saudável

- mucosa da laringe com **aparência homogênea**, ausência de lesões.

### Lesões da Laringe

- A **leucoplasia** da prega vocal laringea (lesão anormal de mucosa, manifestada por placas brancas epiteliais planas ou espessas ou manchas)(PANWAR et al, 2013).
- **Vasos hipertróficos** aferentes (KLIMZA, 2018).
- Lesões em padrão de **alças capilares intrapapilares** na mucosa da laringe (NI et al., 2011).



Tecido saudável



Figura 9. Slide 9

## Referencial Teórico

### Processamento Digital de Imagens (PDI)

- **Imagem Digital (monocromática)** → representação de uma imagem analógica sob forma de uma matriz bidimensional.
  - Constituída de **pixels** (GONZALEZ et al., 2004).
- **Etapas do PDI:**
  - ✓ **Aquisição** → obtenção de imagens em formato digital;
  - ✓ **Pré-processamento** → manipulações na imagem de entrada para melhorar a qualidade;
  - ✓ **Segmentação** → dividir a imagem em porções menores, com foco nos objetos de interesse;
  - ✓ **Extração das características** → captura e transformação de atributos (cor, textura, etc.) em dados quantitativos
  - ✓ **Reconhecimento e interpretação** → classificação, com atribuição de rótulos aos objetos identificados na imagem, conforme suas características, e interpretação dos resultados.
- **Algoritmos de aprendizagem de máquina** tem se tornado um método bastante utilizado para promover reconhecimento de imagens e objetos, com excelentes resultados.

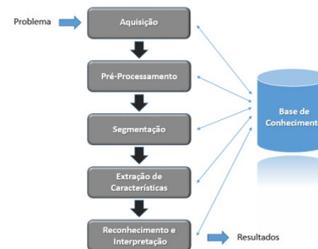


Figura 10. Slide 10

## Referencial Teórico

### Redes Neurais Artificiais e Classificação de Imagens Médicas

**Rede Neural Artificial:** simulação computacional do processo humano de aprendizagem e transmissão de informações, inspirada no funcionamento dos neurônios biológicos.

➢ Possui uma camada de entrada, uma **camada de saída** e pelo menos uma **camada "oculta"** intermediária, em que cada camada está totalmente conectada à seguinte.

➢ algoritmo de **retropropagação:** utiliza de gradientes, e consiste na minimização dos erros de propagação, partindo do erro da saída e retrocedendo esse erro em direção à entrada, atualizando os pesos entre as conexões dos neurônios artificiais.

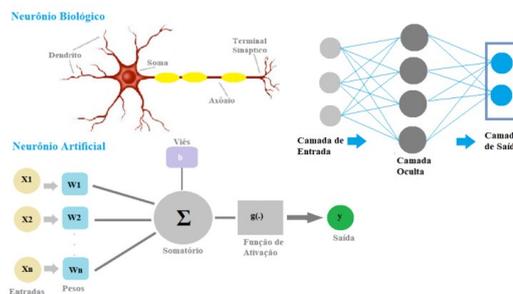


Figura 2.10. Estrutura de um neurônio artificial.



Figura 11. Slide 11

## Referencial Teórico

### Redes Neurais Artificiais e Classificação de Imagens Médicas

#### Redes Neurais Convolucionais (CNN)

- Baseadas em processos biológicos do sistema visual
- Conjunto de camadas distintas de convolução e subamostragem empilhadas
- Recebem um vetor, e transformam em um valor escalar.
- Bom desempenho em tarefas de visão computacional, porém apresentam limitação no reconhecimento da relação entre os objetos e o todo na imagem.

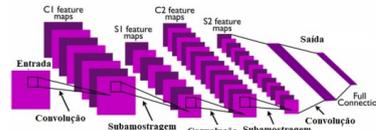


Figura 2.12. Estrutura de uma rede CNN em dois estágios (LeCun et al., 2010).

12

Figura 12. Slide 12

## Referencial Teórico

#### Data Augmentation

- Técnicas que permitam ampliar o conjunto de imagens que serão utilizados no treinamento da rede e melhoram o desempenho.
- Redes Generativas Adversárias (GAN):
  - capaz gerar amostras artificiais a partir de suposições sobre a distribuição dos dados
  - duas redes (geradora e discriminadora) colocadas uma contra a outra:
    - Geradora → gera novas instâncias de dados
    - Discriminadora → avalia sua autenticidade

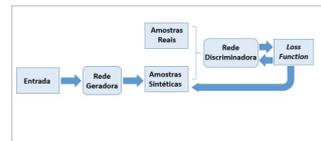


Figura 2.15. Estrutura de uma rede GAN.

Figura 9 – Esquema de uma GAN.  
Fonte: <https://towardsdatascience.com/semi-supervised-learning-and-gans-f23bbf4ac683>

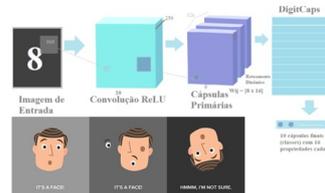
13

Figura 13. Slide 13

## Referencial Teórico

#### Redes Neurais de Cápsula (CapsNet)

- Agrupamento de neurônios artificiais em cápsulas
- Algoritmo de roteamento por concordância: um subgrupo de cápsulas em uma camada “concorda” com a detecção de um padrão, uma cápsula em nível maior é ativada, em seguida outra e assim sucessivamente.
- Parâmetros de pose (orientação, posição, escala, relação de hierarquia entre objetos na imagem)
- Capaz de reconhecer relação de hierarquia entre objetos e o todo na imagem.



Algorithm 1: Dynamic Routing, proposed by Sabour et al. [1]

```

1 Given:  $u_{ji}, r, l, l + 1$ 
2  $b_{i,j} = 0$  for all capsule  $i$  in layer  $l$  and  $j$  in layer  $l + 1$ 
3 for  $i = 0$  to  $r$  do
4   for all capsule  $i$  in layer  $l$ ,  $c_i = \text{Softmax}(b_i)$ 
5   for all capsule  $j$  in layer  $l + 1$ ,  $v_j = \text{Squash}(\sum_i c_i u_{ji})$ 
6   for all capsule  $i$  in layer  $l$  and  $j$  in  $l + 1$ ,  $b_{i,j} = b_{i,j} + u_{ji} v_j$ 
7 end
8 return  $v_j$ 
    
```

Figura 5 – Arquitetura de uma CapsNet.  
Fonte: Huang, W., Zhou, F. DA-CapsNet: dual attention mechanism capsule network. So Rep 10, 11383 (2020). <https://doi.org/10.1038/s41598-020-68453-w>



Figura 14. Slide 14

## Referencial Teórico

### • Trabalhos Relacionados

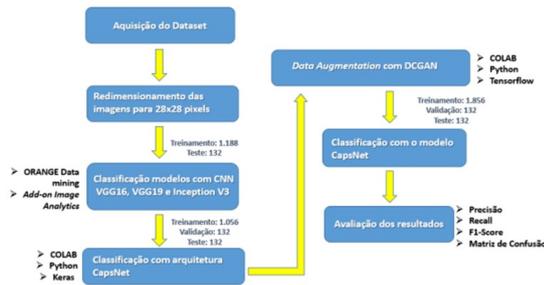
Autor(es)	Tipo de Classificação	Resumo	Recall médio
(Araújo et al., 2019)	Multiclasses	SVM e CNN	98%
(Moccia et al., 2017)	Multiclasses	SVM	98%
(Kwon et al., 2022)	Binária	Árvores de Decisão	97%
(Inaba et al., 2020)	Binária	Retina Net	95,5%
(Azam et al., 2022b)	Multiclasses	SegMENT	95%
(Xiong et al., 2019)	Multiclasses	CNN	92,2%
(He et al., 2021)	Binária	CNN	90,1%
(Esmaili et al., 2020)	Binária	RF	84,6%
(Azam et al., 2022a)	Binária	CNN YOLO	62%

Embora existam pesquisas explorando a classificação de tumores na laringe a partir de CNN, são escassos os estudos relacionados à utilização de classificadores baseados em redes de cápsula para identificação de câncer a partir de imagens do tecido laringeo obtidas por NBI.

15

Figura 15. Slide 15

## Material e métodos



Computador pessoal com processador Intel Core i7, 2.40 GHz e 2 núcleos, com 8 GB de RAM



Figura 9 – Metodologia.

16

Figura 16. Slide 16

## Material e métodos

### Aquisição do Bancos de Imagens:

- 1.320 imagens de tecidos laringeos saudáveis e cancerosos em estágio inicial, obtidas por MOCCLIA et al. (2017)
- 4 classes:
  - TS (tecido saudável);
  - TVH (tecido com vasos hipertróficos);
  - TLE (tecido com leucoplasia);
  - TACI (tecido com alças capilares intrapapilares)
- Redimensionadas para 28x28 pixels

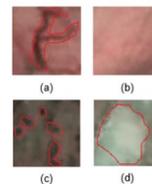


Figura 3.2. Amostra do dataset de imagens do tecido da laringe: (a) tecido com vasos hipertróficos, (b) tecido saudável, (c) tecido com alças capilares intrapapilares, (d) tecido com leucoplasia.

17

Figura 17. Slide 17

## Material e métodos

### Data Augmentation com GAN

- Geração de imagens sintéticas com DCGAN.
- Criação de um banco de imagens ampliado.
- Utilizadas 440 imagens pertencentes ao dataset original (100x100), sendo 110 de cada classe, como entrada.
- Redimensionadas para 28x28.
- O número de épocas de treinamento foi variável para cada classe de imagens.

	Imagens Geradas	Épocas
TVH	200	700
TACI	200	1000
TLE	200	700
TS	200	350

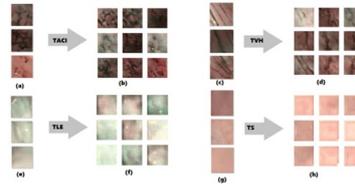


Figura 4.1. Amostra das imagens obtidas a partir da aplicação da técnica de data augmentation, onde (a), (c), (e) e (g) são amostras das imagens do dataset original e (b),(d),(f) e (h) correspondem as imagens geradas artificialmente com DCGAN.

Figura 9 – Metodologia.

18

Figura 18. Slide 18

## Material e métodos

### Classificação:

- CNN VGG 16 → Orange (100 épocas de treino)
- CNN VGG 19 → Orange (100 épocas de treino)
- CNN Inception v3 → Orange (100 épocas de treino)
- CapsNet → Colab (400 épocas de treino)
- CapsNet com Aumento de Dados (DCGAN) → Colab (700 épocas de treino)

Arquitetura	Imagens Treinamento
CNN VGG 16	1.188
CNN VGG 19	1.188
CNN Inception v3	1.188
CapsNet	1.056 + 132 validação
CapsNet + DCGAN	1.856 + 132 validação

132 imagens no conjunto de testes

19

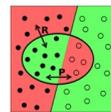
Figura 19. Slide 19

## Material e métodos

### Avaliação dos Resultados:

- Os resultados obtidos pelos classificadores serão avaliados a partir da matriz de confusão, utilizando as seguintes métricas: precisão, recall e F1.

- VP = verdadeiros positivos
- FP = falsos positivos
- VN = verdadeiros negativos
- FN = falsos negativos



$$Precis\tilde{a}o = \frac{VP}{VP + FP}$$

$$Recall = \frac{VP}{VP + FN}$$

$$F1 - Score = \frac{2 * Precis\tilde{a}o * Recall}{Precis\tilde{a}o + Recall}$$

20

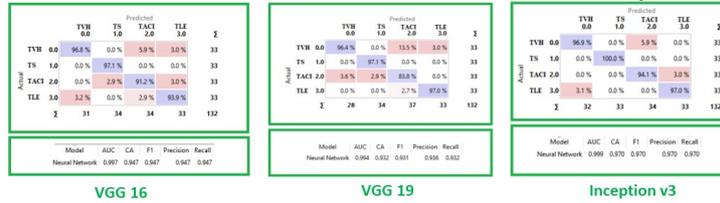
Figura 20. Slide 20

# Resultados e Discussão

## Algoritmos CNN:

Classes:

0.0 - Vasos Hipertrofos 2.0 - Alças Capilares Intrapapilares  
1.0 - Tecido Saudável 3.0 - Leucoplasia



21

Figura 21. Slide 21

# Resultados e Discussão

Tabela 4.2. Recall na classificação com CapsNet, por classe

Classe	Dataset Original	Dataset Aumentado com GAN
Vasos Hipertrofos	0.788	0.879
Tecido Saudável	1.000	1.000
Alças Capilares Intrapapilares	0.939	0.969
Leucoplasia	1.000	1.000

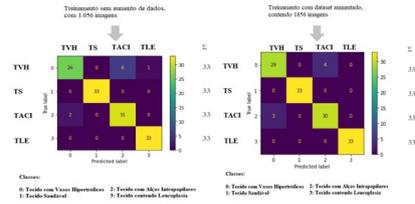
Tabela 4.3. Precisão na classificação com CapsNet, por classe

Classe	Dataset Original	Dataset Aumentado com GAN
Vasos Hipertrofos	0.928	0.906
Tecido Saudável	1.000	1.000
Alças Capilares Intrapapilares	0.838	0.882
Leucoplasia	0.971	1.000

Tabela 4.4. F1-Score na classificação com CapsNet, por classe

Classe	Dataset Original	Dataset Aumentado com GAN
Vasos Hipertrofos	0.852	0.892
Tecido Saudável	1.000	1.000
Alças Capilares Intrapapilares	0.886	0.895
Leucoplasia	0.985	1.000

Arquitetura	F1-score	Precisão	Recall
CapsNet	0.931	0.934	0.932
CapsNet + DCGAN	0.947	0.947	0.947



22

Figura 22. Slide 22

# Resultados e Discussão

Tabela 4.5. Resumo dos Resultados

Arquitetura	F1-score	Precisão	Recall	Imagens Treinamento
CNN VGG 16	0.947	0.947	0.947	1.188
CNN VGG 19	0.931	0.936	0.932	1.188
CNN Inception v3	0.970	0.970	0.970	1.188
CapsNet	0.931	0.934	0.932	1.056 + 132 validação
CapsNet + DCGAN	0.947	0.947	0.947	1.856 + 132 validação

Quadro 4.1. Análise dos resultados junto aos trabalhos relacionados

Autores)	Tipo de Classificação	Resumo	Recall médio
(Arújo et al., 2019)	Multiclasses	SVM e CNN	98%
(Moccia et al., 2017)	Multiclasses	SVM	98%
Costa, 2023*	Multiclasses	CNN Inception V3 utilizando programação visual com Orange Data Mining	97%
(Kwon et al., 2022)	Binária	Árvores de Decisão	97%
(Inaba et al., 2020)	Binária	Retina Net	95,5%
(Azam et al., 2022b)	Multiclasses	SEGMENT	95%
Costa, 2023*	Multiclasses	CapsNet e DCGAN	94,7%
(Xiong et al., 2019)	Multiclasses	CNN	92,2%
(He et al., 2021)	Binária	CNN	90,1%
(Esmell et al., 2020)	Binária	RF	84,6%
(Azam et al., 2022a)	Binária	CNN YOLO	62%

Fonte: elaboração própria.

23

Figura 23. Slide 23

## Conclusão

- ✓ O classificador baseado em **rede de cápsulas** aliado ao recurso de **aumento de dados com DCGAN**, apresentou desempenho satisfatório em classificar lesões da laringe em imagens obtidas por NBI, com **recall, precisão e F1-Score médios de 94,7%**.
- ✓ A implementação dos classificadores CNN com **programação visual** no software livre Orange permitiu o alcance de ótimos resultados, tendo por vantagem a **facilidade de utilização** e por **não exigir conhecimentos avançados de programação**.
- ✓ O modelo CNN baseado em **Inception V3**, implementado com programação visual no software livre Orange, obteve o **melhor desempenho geral**, permitindo a classificação das imagens com recall, precisão e F1-Score médios equivalentes a **97%**, próximos ao estado da arte (98%).
- ✓ O **acréscimo de imagens sintéticas** geradas por GAN influenciou **positivamente** o desempenho do classificador baseado em rede de cápsula.



Figura 24. Slide 24

## Conclusão

- **Limitações:**
  - Aplicação da técnica de aumento da base de dados com GAN utilizando apenas um modelo baseado em rede de cápsula.
  - Baixa resolução das imagens (tamanho de 28 pixels por 28 pixels).
- **Trabalhos Futuros:**
  - Realizar o treinamento e classificação com imagens em **maior resolução**;
  - Gerar maior um **maior número de imagens sintéticas** com GAN;
  - Explorar o modelo CapsNet com aumento de dados por GAN na classificação de **outras patologias em imagens médicas**;
  - Aplicar **aumento de dados com GAN em outros modelos de classificadores**;
  - Integrar o modelo de inteligência artificial CapsNet e GAN em uma **interface gráfica de usuário**.



Figura 25. Slide 25