



Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ciência da Computação

Uso de Few-shot Learning para lidar com o Desbalanceamento Severo de Dados na Classificação de Lesões Glomerulares

Alexsandro Silva Santos

Feira de Santana

2025



Universidade Estadual de Feira de Santana
Programa de Pós-Graduação em Ciência da Computação

Alexsandro Silva Santos

**Uso de Few-shot Learning para lidar com o
Desbalanceamento Severo de Dados na Classificação de
Lesões Glomerulares**

Dissertação apresentada à Universidade
Estadual de Feira de Santana como parte
dos requisitos para a obtenção do título de
Mestre em Ciência da Computação.

Orientador: Angelo Amâncio Duarte

Feira de Santana

2025

Ficha Catalográfica – Biblioteca Central Julieta Carteadó

S233u Santos, Alexsandro Silva

Uso de Few-shot Learning para lidar com o desbalanceamento severo de dados na classificação de lesões glomerulares./ Alexsandro Silva Santos. – 2025.
200 f.: il.

Orientador: Angelo Amâncio Duarte

Dissertação (mestrado) – Universidade Estadual de Feira de Santana, Programa de Pós-Graduação em Ciência da Computação, 2025.

1.Patologia Computacional. 2.Desbalanceamento de classes. 3.Few-shot Learning. I.Duarte, Angelo Amâncio, orient. II.Universidade Estadual de Feira de Santana. III.Título.

CDU : 004:616.61

Maria de Fátima de Jesus Moreira – Bibliotecária – CRB5/1120

Alexsandro Silva Santos

Uso de Few-shot Learning para lidar com o Desbalanceamento Severo de Dados na Classificação de Lesões Glomerulares

Dissertação apresentada à Universidade Estadual de Feira de Santana como parte dos requisitos para a obtenção do título de Mestre em Ciência da Computação.

Feira de Santana, 4 de julho de 2025

BANCA EXAMINADORA

Documento assinado digitalmente
 ANGELO AMANCIO DUARTE
Data: 07/07/2025 14:57:06-0300
Verifique em <https://validar.iti.gov.br>

Angelo Amâncio Duarte (Orientador)
Universidade Estadual de Feira de Santana

Documento assinado digitalmente
 ROBERTA BARBOSA OLIVEIRA
Data: 06/07/2025 23:19:17-0300
Verifique em <https://validar.iti.gov.br>

Roberta Barbosa Oliveira
Universidade de Brasília

Documento assinado digitalmente
 MATHEUS GIOVANNI PIRES
Data: 07/07/2025 10:47:24-0300
Verifique em <https://validar.iti.gov.br>

Matheus Giovani Pires
Universidade Estadual de Feira de Santana

Abstract

In the field of nephropathology, the automatic classification of renal biopsy images is crucial for supporting patient diagnosis and treatment. Machine Learning (ML) techniques—particularly those based on Convolutional Neural Networks (CNNs)—have enabled the automation of this task, offering operational improvements. However, these techniques typically require large, balanced datasets to achieve strong performance and are negatively impacted in scenarios with limited data and class imbalance—conditions commonly found in the classification of rare renal lesions. This study presents an applied investigation of approaches for handling class imbalance, including well-established techniques from the literature and the use of Few-Shot Learning (FSL). The results demonstrate that it is possible to achieve F1-scores above 85% across different imbalance scenarios encountered in the analyzed lesion types, reaching scores above 90% even in cases with imbalance ratios as high as $\approx 1:30$. Nevertheless, the experiments conducted in this study revealed that data imbalance, when considered in isolation, is not the sole factor influencing the final performance of ML models.

Keywords: Computational Pathology, Class Imbalance, Few-shot Learning.

Resumo

No contexto da nefropatologia, a classificação automática de imagens de biópsias renais é essencial para apoiar o diagnóstico e tratamento de pacientes. Técnicas de Machine Learning (ML), especialmente aquelas baseadas em redes neurais convolucionais (Convolutional Neural Networks – CNN), têm permitido a automatização dessa tarefa, oferecendo ganhos operacionais. No entanto, essas técnicas geralmente requerem grandes volumes de dados balanceados para alcançar bom desempenho, sendo negativamente afetadas em cenários com escassez de dados e desbalanceamento de classes — uma condição comum na classificação de lesões renais raras. Este trabalho apresenta um estudo aplicado de abordagens para lidar com o desbalanceamento de classes, incluindo o uso de técnicas consolidadas na literatura e a aplicação do aprendizado com poucas amostras (Few-Shot Learning – FSL). Os resultados demonstram que é possível atingir desempenho superior a 85% de F1-score em diferentes cenários de desbalanceamento, encontrados para as lesões aqui analisadas, chegando a números superiores a 90% mesmo com taxas de desbalanceamento da ordem de $\approx 1:30$. No entanto, os experimentos realizados neste estudo evidenciaram que o desbalanceamento de dados, considerado isoladamente, não é o único fator relevante no desempenho final dos modelos de ML.

Palavras-chave: Patologia Computacional, Desbalanceamento de classes, Few-shot Learning

Prefácio

Esta dissertação de mestrado foi submetida à Universidade Estadual de Feira de Santana (UEFS) como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

A dissertação foi desenvolvida no Programa de Pós-Graduação em Ciência da Computação (PGCC), tendo como orientador o Prof. Dr. Angelo Amâncio Duarte.

Esta pesquisa foi financiada pelo programa FINAPESQ UEFS termo de outorga 074/2021.

Agradecimentos

Agradeço a Deus, pelo milagre da vida, pelo cuidado para comigo e pela oportunidade de estar aqui; à minha família e, em especial, à minha amada esposa, Vanny, por ser minha companheira de jornada e pela paciência e suporte durante os últimos meses; à minha filha, Rayssa, cuja admiração e respeito me motivam a continuar o meu caminho de aperfeiçoamento.

Agradeço à minha querida e amada mãe, a D. Helena, que sempre me valorizou muito mais do que eu merecia e sempre me amou e ama incondicionalmente. Manifesto meu sincero agradecimento aos familiares e amigos, que torceram por mim e entenderam esse momento da minha vida, em que me encontrei com menos tempo e, às vezes, com menos paciência.

Agradeço ao Instituto Federal de Educação Ciência e Tecnologia Baiano, sem o qual a minha presença neste programa não seria possível.

Preciso, deveras, manifestar o meu profundo agradecimento ao professor Angelo Amâncio Duarte, cujo rigor, paciência e comprometimento tornaram possível este trabalho.

Agradeço aos professores, servidores, funcionários, colegas e, por que não dizer, amigos, que encontrei na Universidade Estadual de Feira de Santana, que deixaram um pouco de si comigo, fazendo-me, indubitavelmente, um ser humano melhor do que eu era antes de ingressar no Programa de Pós-graduação em Ciências da Computação. A todos, faço os melhores votos de paz, saúde, sucesso e prosperidade.

*“Com a sabedoria se edifica a casa,
e com o entendimento ela se
estabelece; e pelo conhecimento se
encherão as câmaras com todos os
bens preciosos e agradáveis.”*

*(A Bíblia Sagrada, Provérbios
24:3-4)*

Sumário

Abstract	i
Resumo	ii
Prefácio	iii
Agradecimentos	iv
Alinhamento com a Linha de Pesquisa	xi
Produções Bibliográficas, Produções Técnicas e Premiações	xii
Lista de Tabelas	xxii
Lista de Figuras	xxvi
Lista de Abreviações	xxvii
Lista de Símbolos	xxix
1 Introdução	1
1.1 Objetivos	4
1.1.1 Objetivo Principal	4
1.1.2 Objetivos Específicos	4
1.2 Contribuições	4
1.3 Organização do Trabalho	5
2 Fundamentação Teórica	6
2.1 Conceitos de Nefropatologia	6
2.1.1 Rins e Glomérulo	6
2.1.2 Biópsia	8
2.1.3 Lesões Glomerulares	10
2.2 Conceitos de Computação	11
2.2.1 Aprendizado de Máquina	11
2.2.2 Visão Computacional	11

2.2.3	Redes Neurais Convolucionais	12
2.2.4	Desbalanceamento de Classes	12
2.2.5	Complexidade dos Dados	13
2.2.6	Abordagens mais comuns para tratar o Desbalanceamento de Classes	15
2.2.7	Métricas de Desempenho em Cenários de Desbalanceamento .	17
2.2.8	Few-shot Learning	20
2.2.9	Few-shot Learning como estratégia para tratar o Desbalance- amento de Classes	23
2.3	Análise Estatística	24
3	Trabalhos Relacionados	26
3.1	Trabalhos que utilizaram Técnicas Comuns para tratar o Desbalan- ceamento de Classes	26
3.2	Trabalhos que utilizaram Few-shot Learning e CIFSL	31
3.3	Análise dos Trabalhos Relacionados	33
3.3.1	Análise dos trabalhos com Técnicas Comuns	33
3.3.2	Análise dos trabalhos que fizeram uso de FSL e CIFSL	34
4	Metodologia	39
4.1	O Conjunto de Dados Original	40
4.2	O Conjunto de Dados Refinado	43
4.3	Ambiente Computacional	49
4.4	Organização dos Conjuntos de Dados para os Experimentos	50
4.5	Modelos Convencionais Usando CNN	51
4.6	Modelos Few-shot Learning	52
4.7	Modelos Ensemble-based	54
4.8	Análise da Complexidade dos Dados	54
4.9	Comparação Estatística dos Modelos	54
5	Resultados e Discussão	56
5.1	Resultados e Análise para a Classe Amiloidose	56
5.2	Resultados e Análise para a Classe Normal	61
5.3	Resultados e Análise para a Classe Esclerose Pura sem Crescente . .	65
5.4	Resultados e Análise para a Classe Hiper celularidade	69
5.5	Resultados e Análise para a Classe Hiper celularidade Pura sem Cres- cente	73
5.6	Resultados e Análise para a Classe Glomerulonefrite Crescente	76
5.7	Resultados e Análise para a Classe Nefropatia Membranosa	80
5.8	Resultados e Análise para a Classe Esclerose	84
5.9	Resultados e Análise para a Classe Podocitopatia	89
5.10	Análise dos Resultados	93
5.10.1	Comparação entre FSL e Métodos Comuns	94
5.10.2	Influência da Curadoria do Dataset	96

5.10.3	Influência dos Corantes	97
5.10.4	Particularidades das Classes Esclerose Pura e Hipercelularidade Pura	97
5.10.5	Comportamento das Métricas Precisão e Revocação (Precision e Recall)	98
5.10.6	Complexidade dos Conjuntos	98
6	Conclusões	99
6.1	Trabalhos Futuros	101
	Referências	103
A	Descrição dos Conjuntos de Dados Utilizados nos Experimentos	112
B	Pesos Utilizados para o Aprendizado Sensível ao Custo	117
C	Relação Completa dos Resultados Alcançados	120
C.1	Resultados para os Modelos Treinados com Métodos Comuns sobre Conjuntos sem Curadoria e sem Rebalanceamento	120
C.1.1	Resultados para os Modelos Treinados com Conjuntos sem Curadoria e sem Restrição por Corante	120
C.1.2	Resultados para os Modelos Treinados com Conjuntos sem Curadoria e Restritos ao Corante HE	122
C.1.3	Resultados para os Modelos Treinados com Conjuntos sem Curadoria e Restritos ao Corante PAS	124
C.2	Resultados dos Modelos Treinados sobre os Conjuntos Curados e sem Superamostragem	126
C.2.1	Resultados para os Modelos Treinados com Conjuntos com Curadoria e sem Restrição de Corante	126
C.2.2	Resultados para os Modelos Treinados com Conjuntos com Curadoria e Restritos ao Corante HE	128
C.2.3	Resultados para os Modelos Treinados com Conjuntos com Curadoria e Restritos ao Corante PAS	129
C.3	Resultados dos Modelos Treinados sobre os Conjuntos sem Curadoria e com Superamostragem	131
C.3.1	Resultados Obtidos sobre os Conjuntos sem Curadoria e sem Restrição por Corante	131
C.3.2	Resultados dos Experimentos com Métodos Comuns sobre os conjuntos HE sem Curadoria e com Superamostragem	133
C.3.3	Resultados dos Experimentos com Métodos Comuns sobre os conjuntos PAS sem Curadoria e com Superamostragem	135
C.4	Resultados para os Modelos treinados com os conjuntos Curados e com Superamostragem	136
C.4.1	Resultados para os Modelos Treinados com Conjuntos Curados e sem Restrição por Corante	137

C.4.2	Resultados para os Modelos treinados com os conjuntos Curados Restritos ao Corante HE e com Superamostragem	138
C.4.3	Resultados para os Modelos treinados com os conjuntos Curados Restritos ao Corante PAS e com Superamostragem	140
C.5	Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos sem Curadoria	142
C.5.1	Resultados para os Modelos Treinados com os Conjuntos sem Curadoria e sem Restrição por Corante	142
C.5.2	Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos sem Curadoria Restritos ao Corante HE	144
C.5.3	Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Sem Curadoria Restritos ao Corante PAS	146
C.6	Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Curados	147
C.6.1	Resultados para os Modelos Treinados com Conjuntos Curados e sem Restrição por Corante	148
C.6.2	Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Curados Restritos ao Corante HE	149
C.6.3	Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Curados Restritos ao Corante PAS	151
D	Gráficos dos Modelos de Maior Desempenho	154
D.1	Amiloidose	154
D.1.1	Curvas da Função de Custo	154
D.1.2	Curvas precision-recall	155
D.1.3	Matrizes de Confusão	155
D.2	Normal	156
D.2.1	Curvas da Função de Custo	156
D.2.2	Curvas precision-recall	156
D.2.3	Matrizes de Confusão	157
D.3	Esclerose Pura sem Crescente	157
D.3.1	Curvas da Função de Custo	157
D.3.2	Curvas precision-recall	158
D.3.3	Matrizes de Confusão	158
D.4	Hipercelularidade	159
D.4.1	Curvas da Função de Custo	159
D.4.2	Curvas precision-recall	159
D.4.3	Matrizes de Confusão	160
D.5	Hipercelularidade Pura sem Crescente	160
D.5.1	Curvas da Função de Custo	160

D.5.2	Curvas precision-recall	161
D.5.3	Matrizes de Confusão	161
D.6	Glomerulonefrite Crescente	162
D.6.1	Curvas da Função de Custo	162
D.6.2	Curvas precision-recall	162
D.6.3	Matrizes de Confusão	163
D.7	Nefropatia Membranosa	163
D.7.1	Curvas da Função de Custo	163
D.7.2	Curvas precision-recall	164
D.7.3	Matrizes de Confusão	164
D.8	Esclerose	165
D.8.1	Curvas da Função de Custo	165
D.8.2	Curvas precision-recall	165
D.8.3	Matrizes de Confusão	166
D.9	Podocitopatía	166
D.9.1	Curvas precision-recall	167
D.9.2	Matrizes de Confusão	167

Alinhamento com a Linha de Pesquisa

Linha de Pesquisa: Computação Inteligente

Este trabalho realiza um estudo sistemático e planejado sobre uma base de dados para a extração de conhecimento e descoberta de padrões, com o uso de algoritmos computacionais, estando adequado com a linha de pesquisa *Computação Inteligente*.

Produções Bibliográficas, Produções Técnicas e Premiações

SANTOS, A. S.; OLIVEIRA, L. R.; DOS-SANTOS, W. L. C.; DUARTE, A. A.. Addressing Class Imbalance in Renal Amyloidosis Classification: A Comparative Study of Few-Shot Learning and Conventional Machine Learning Techniques In: 5th International Conference on Image Processing and Vision Engineering (IMPROVE 2025), 2025, Porto, Lisbon. Proceedings of the 5th International Conference on Image Processing and Vision Engineering (IMPROVE 2025). Berlin, Germany: Springer Nature, 2025.

Lista de Tabelas

2.1	Matriz de Confusão para um problema de classificação binário. . . .	17
3.1	Sumário das técnicas e resultados para os trabalhos realizados na patologia computacional	30
3.2	Sumário das técnicas e resultados para os trabalhos relacionados a terem empregado métodos comuns fora da patologia computacional .	31
3.3	Sumário dos resultados para FSL na patologia computacional. . . .	33
3.4	Sumário dos resultados para FSL fora da patologia computacional. .	33
4.1	Distribuição do conjunto de dados original por lesão e corante. . . .	40
4.2	Distribuição do conjunto de dados original por lesão e corante em valores percentuais.	41
4.3	Distorções encontradas no conjunto.	45
4.4	Percentual de imagens contendo glomérulos parcialmente representados, separado por classe.	48
4.5	Distribuição do conjunto de dados curado por lesão e corante. . . .	48
4.6	Distribuição do conjunto de dados curado, por lesão e corante em valores percentuais.	49
5.1	Medidas de complexidade para a classe <i>amiloidose</i>	58
5.2	Taxas mais altas de F1-score alcançadas para <i>amiloidose</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas ResNet-50, EfficientNet-B0 e Inception-v3, sobre o conjunto HE.	58
5.3	Medidas de complexidade para a classe <i>normal</i>	62
5.4	Taxas mais altas de F1-score alcançadas para <i>normal</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas Inception-v3, VGG-16 e ResNet-18, sobre o conjunto ROS-HE. .	62
5.5	Medidas de complexidade para a classe <i>esclerose pura sem crescente</i> . .	66
5.6	Taxas mais altas de F1-score alcançadas para <i>esclerose pura sem crescente</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas KCN, VGG-16 e EfficientNet-B0, sobre o conjunto <i>Baseline</i>	66
5.7	Medidas de complexidade para a classe <i>hipercelularidade</i>	70

5.8	Taxas mais altas de F1-score alcançadas para <i>hipercelularidade</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas KCN, VGG-16 e EfficientNet-B0, sobre o conjunto ROS-Curated-HE.	70
5.9	Medidas de complexidade para a classe <i>hipercelularidade pura sem crescente</i>	74
5.10	Taxas mais altas de F1-score alcançadas para <i>hipercelularidade pura sem crescente</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas VGG-16, Inception-v3 e ResNet-18, sobre o conjunto ROS-PAS.	74
5.11	Medidas de complexidade para a classe <i>glomerulonefrite crescente</i> . . .	78
5.12	Taxas mais altas de F1-score alcançadas para <i>glomerulonefrite crescente</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas VGG-16, Inception-v3 e ResNet-18, sobre o conjunto <i>Curated-glomerulonefrite</i>	78
5.13	Medidas de complexidade para a classe <i>nefropatia membranosa</i>	81
5.14	Taxas mais altas de F1-score alcançadas para <i>nefropatia membranosa</i> . O modelo <i>ensemble-based</i> foi construído pelos modelos produzidos pelas arquiteturas Inception-v3, ResNet-50 e ResNet-18, sobre o conjunto <i>Curated-HE-Membranosa</i>	82
5.15	Medidas de complexidade para a classe <i>esclerose</i>	85
5.16	Taxas mais altas de F1-score alcançadas para <i>esclerose</i> . O modelo <i>ensemble-based</i> foi construído com modelos produzidos pelas arquiteturas ResNet-50, ResNet-18 e Inception-v3, sobre o conjunto <i>PAS-esclerose</i>	86
5.17	Medidas de complexidade para a classe <i>podocitopatia</i> . O modelo <i>ensemble-based</i> foi construído com os modelos produzidos pelas arquiteturas ResNet-50, ResNet-18 e Inception-v3, sobre o conjunto <i>PAS-podocitopatia</i>	90
5.18	Taxas mais altas de F1-score alcançadas para <i>podocitopatia</i>	90
5.19	Resultados mais altos obtidos para cada classe com modelos individuais.	93
5.20	Resultados mais altos obtidos para cada classe com modelos <i>ensemble-based</i> . A coluna <i>Classif.</i> aponta o tipo de modelo a ter apresentado o maior resultado.	94
5.21	Resultados do teste de Wilcoxon.	95
5.22	Resultados obtidos para a classe <i>hipercelularidade</i> com diferentes tamanhos para o conjunto de suporte (apenas modelos individuais). Os valores apresentados para o teste de Wilcoxon são referentes à comparação com modelos treinados com <i>200-shot</i>	95
5.23	Resultados obtidos para as classes <i>amiloidose</i> , <i>esclerose</i> e <i>podocitopatia</i> com diferentes tamanhos para o conjunto de suporte (apenas modelos individuais). Os valores apresentados para o teste de Wilcoxon são referentes à comparação com modelos treinados com <i>(25%)-shot</i> . .	96

5.24	Comparação entre o uso da <i>similaridade do cosseno</i> ($\text{sim}_{\cos}(\mathbf{x}, \mathbf{y})$) e a <i>distância euclidiana</i> ($\ \mathbf{x} - \mathbf{y}\ _2$).	96
5.25	Resultados do teste de Wilcoxon. Comparação entre modelos treinados com conjuntos monocorante (Mono) e multicorante (Multi). . .	97
A.1	Descrição dos conjuntos de dados criados para este trabalho.	112
B.1	Pesos utilizados para os conjuntos <i>Baseline-[nome da classe]</i> e <i>ROS-[nome da classe]</i>	117
B.2	Pesos utilizados para os conjuntos <i>HE-[nome da classe]</i> e <i>ROS-HE-[nome da classe]</i>	117
B.3	Pesos utilizados para os conjuntos <i>PAS-[nome da classe]</i> e <i>ROS-PAS-[nome da classe]</i>	118
B.4	Pesos utilizados para os conjuntos <i>Curated-[nome da classe]</i> e <i>ROS-Curated-[nome da classe]</i>	118
B.5	Pesos utilizados para os conjuntos <i>Curated-HE-[nome da classe]</i> e <i>ROS-Curated-HE-[nome da classe]</i>	118
B.6	Pesos utilizados para os conjuntos <i>Curated-PAS-[nome da classe]</i> e <i>ROS-Curated-PAS-[nome da classe]</i>	119
C.1	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-amiloidose</i>	120
C.2	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-normal</i>	121
C.3	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-esclerose-pura</i>	121
C.4	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-hipercelularidade</i>	121
C.5	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-hipercelularidade-pura</i>	121
C.6	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-glomerulonefrite</i>	121
C.7	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-membranosa</i>	122
C.8	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-esclerose</i>	122
C.9	Resultados dos experimentos realizados sobre o conjunto <i>Baseline-podocitopatia</i>	122
C.10	Resultados dos experimentos realizados sobre o conjunto <i>HE-amiloidose</i>	122
C.11	Resultados dos experimentos realizados sobre o conjunto <i>HE-normal</i>	122
C.12	Resultados dos experimentos realizados sobre o conjunto <i>HE-esclerose-pura</i>	123
C.13	Resultados dos experimentos realizados sobre o conjunto <i>HE-hipercelularidade</i>	123

C.14 Resultados dos experimentos realizados sobre o conjunto <i>HE-glomerulonefrite</i>	123
C.15 Resultados dos experimentos realizados sobre o conjunto <i>HE-membranosa</i>	123
C.16 Resultados dos experimentos realizados sobre o conjunto <i>HE-esclerose</i>	123
C.17 Resultados dos experimentos realizados sobre o conjunto <i>HE-podocitopatia</i>	124
C.18 Resultados dos experimentos realizados sobre o conjunto <i>PAS-amiloidose</i>	124
C.19 Resultados dos experimentos realizados sobre o conjunto <i>PAS-normal</i>	124
C.20 Resultados dos experimentos realizados sobre o conjunto <i>PAS-esclerose-pura</i>	124
C.21 Resultados dos experimentos realizados sobre o conjunto <i>PAS-hipercelularidade</i>	124
C.22 Resultados dos experimentos realizados sobre o conjunto <i>PAS-hipercelularidade-pura</i>	125
C.23 Resultados dos experimentos realizados sobre o conjunto <i>PAS-glomerulonefrite</i>	125
C.24 Resultados dos experimentos realizados sobre o conjunto <i>PAS-membranosa</i>	125
C.25 Resultados dos experimentos realizados sobre o conjunto <i>PAS-esclerose</i>	125
C.26 Resultados dos experimentos realizados sobre o conjunto <i>PAS-podocitopatia</i>	125
C.27 Resultados dos experimentos realizados sobre o conjunto <i>Curated-amiloidose</i>	126
C.28 Resultados dos experimentos realizados sobre o conjunto <i>Curated-normal</i>	126
C.29 Resultados dos experimentos realizados sobre o conjunto <i>Curated-esclerose-pura</i>	126
C.30 Resultados dos experimentos realizados sobre o conjunto <i>Curated-hipercelularidade</i>	126
C.31 Resultados dos experimentos realizados sobre o conjunto <i>Curated-hipercelularidade-pura</i>	127
C.32 Resultados dos experimentos realizados sobre o conjunto <i>Curated-glomerulonefrite</i>	127
C.33 Resultados dos experimentos realizados sobre o conjunto <i>Curated-membranosa</i>	127
C.34 Resultados dos experimentos realizados sobre o conjunto <i>Curated-esclerose</i>	127
C.35 Resultados dos experimentos realizados sobre o conjunto <i>Curated-podocitopatia</i>	127
C.36 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-amiloidose</i>	128

C.37 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-normal</i>	128
C.38 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-esclerose-pura</i>	128
C.39 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-hiper celularidade</i>	128
C.40 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-glomerulonefrite</i>	128
C.41 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-membranosa</i>	129
C.42 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-esclerose</i>	129
C.43 Resultados dos experimentos realizados sobre o conjunto <i>Curated-HE-podocitopatía</i>	129
C.44 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-amiloidose</i>	129
C.45 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-normal</i>	129
C.46 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-esclerose-pura</i>	130
C.47 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-hiper celularidade</i>	130
C.48 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-hiper celularidade-pura</i>	130
C.49 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-glomerulonefrite</i>	130
C.50 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-membranosa</i>	130
C.51 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-esclerose</i>	131
C.52 Resultados dos experimentos realizados sobre o conjunto <i>Curated-PAS-podocitopatía</i>	131
C.53 Resultados dos experimentos realizados sobre o conjunto <i>ROS-amiloidose</i>	131
C.54 Resultados dos experimentos realizados sobre o conjunto <i>ROS-normal</i>	131
C.55 Resultados dos experimentos realizados sobre o conjunto <i>ROS-esclerose-pura</i>	132
C.56 Resultados dos experimentos realizados sobre o conjunto <i>ROS-hiper celularidade</i>	132
C.57 Resultados dos experimentos realizados sobre o conjunto <i>ROS-hiper celularidade-pura</i>	132
C.58 Resultados dos experimentos realizados sobre o conjunto <i>ROS-glomerulonefrite</i>	132

C.59 Resultados dos experimentos realizados sobre o conjunto <i>ROS-membranosa</i>	132
C.60 Resultados dos experimentos realizados sobre o conjunto <i>ROS-esclerose</i> .	133
C.61 Resultados dos experimentos realizados sobre o conjunto <i>ROS-podocitopatia</i>	133
C.62 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-amiloidose</i>	133
C.63 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-normal</i>	133
C.64 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-esclerose-pura</i>	133
C.65 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-hipercelularidade</i>	134
C.66 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-glomerulonefrite</i>	134
C.67 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-membranosa</i>	134
C.68 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-esclerose</i>	134
C.69 Resultados dos experimentos realizados sobre o conjunto <i>ROS-HE-podocitopatia</i>	134
C.70 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-amiloidose</i>	135
C.71 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-normal</i>	135
C.72 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-esclerose-pura</i>	135
C.73 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-hipercelularidade</i>	135
C.74 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-hipercelularidade-pura</i>	135
C.75 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-glomerulonefrite</i>	136
C.76 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-membranosa</i>	136
C.77 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-esclerose</i>	136
C.78 Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-podocitopatia</i>	136
C.79 Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-amiloidose</i>	137
C.80 Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-normal</i>	137

C.81	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-esclerose-pura</i>	137
C.82	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-hiper celularidade</i>	137
C.83	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-hiper celularidade-pura</i>	137
C.84	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-glomerulonefrite</i>	138
C.85	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-membranosa</i>	138
C.86	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-esclerose</i>	138
C.87	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-podocitopatia</i>	138
C.88	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-amiloidose</i>	138
C.89	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-normal</i>	139
C.90	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-esclerose-pura</i>	139
C.91	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-hiper celularidade</i>	139
C.92	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-glomerulonefrite</i>	139
C.93	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-membranosa</i>	139
C.94	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-esclerose</i>	140
C.95	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-HE-podocitopatia</i>	140
C.96	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-amiloidose</i>	140
C.97	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-normal</i>	140
C.98	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-PAS-esclerose-pura</i>	140
C.99	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-PAS-hiper celularidade</i>	141
C.100	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-PAS-hiper celularidade-pura</i>	141
C.101	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-PAS-glomerulonefrite</i>	141
C.102	Resultados dos experimentos realizados sobre o conjunto <i>ROS-PAS-Curated-PAS-membranosa</i>	141

C.103	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-PAS-esclerose</i>	141
C.104	Resultados dos experimentos realizados sobre o conjunto <i>ROS-Curated-PAS-podocitopatia</i>	142
C.105	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-amiloidose</i>	142
C.106	Resultados dos experimentos com FLS realizados sobre o conjunto <i>Baseline-normal</i>	142
C.107	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-esclerose-pura</i>	143
C.108	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-hiper celularidade</i>	143
C.109	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-hiper celularidade-pura</i>	143
C.110	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-glomerulonefrite</i>	143
C.111	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-membranosa</i>	143
C.112	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-esclerose</i>	144
C.113	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Baseline-podocitopatia</i>	144
C.114	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-amiloidose</i>	144
C.115	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-normal</i>	144
C.116	Resultados dos experimentos com FSL realizados sobre o conjunto <i>FSL-HE-esclerose-pura</i>	144
C.117	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-hiper celularidade</i>	145
C.118	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-glomerulonefrite</i>	145
C.119	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-membranosa</i>	145
C.120	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-esclerose</i>	145
C.121	Resultados dos experimentos com FSL realizados sobre o conjunto <i>HE-podocitopatia</i>	145
C.122	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-amiloidose</i>	146
C.123	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-normal</i>	146
C.124	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-esclerose-pura</i>	146

C.125	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-hiper celularidade.</i>	146
C.126	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-hiper celularidade-pura.</i>	146
C.127	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-glomerulonefrite.</i>	147
C.128	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-membranosa.</i>	147
C.129	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-esclerose.</i>	147
C.130	Resultados dos experimentos com FSL realizados sobre o conjunto <i>PAS-podocitopatia.</i>	147
C.131	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-amiloidose.</i>	148
C.132	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-normal.</i>	148
C.133	Resultados dos experimentos realizados sobre o conjunto <i>Curated-esclerose-pura.</i>	148
C.134	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-hiper celularidade.</i>	148
C.135	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-hiper celularidade-pura.</i>	148
C.136	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-glomerulonefrite.</i>	149
C.137	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-membranosa.</i>	149
C.138	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-esclerose.</i>	149
C.139	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-podocitopatia.</i>	149
C.140	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-amiloidose.</i>	149
C.141	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-normal.</i>	150
C.142	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-esclerose-pura.</i>	150
C.143	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-hiper celularidade.</i>	150
C.144	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-glomerulonefrite.</i>	150
C.145	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-membranosa.</i>	150
C.146	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-esclerose.</i>	151

C.147	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-HE-podocitopatia.</i>	151
C.148	Resultados dos experimentos realizados sobre o conjunto <i>Curated-</i> <i>PAS-amiloidose.</i>	151
C.149	Resultados dos experimentos realizados sobre o conjunto <i>Curated-</i> <i>PAS-normal.</i>	151
C.150	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-PAS-esclerose-pura.</i>	151
C.151	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-PAS-hiper celularidade.</i>	152
C.152	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-PAS-hiper celularidade-pura.</i>	152
C.153	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-PAS-glomerulonefrite.</i>	152
C.154	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-PAS-membranosa.</i>	152
C.155	Resultados dos experimentos com FSL realizados sobre o conjunto <i>Curated-PAS-esclerose.</i>	152
C.156	Resultados dos experimentos FSL realizados sobre o conjunto <i>Curated-PAS-podocitopatia.</i>	153

Lista de Figuras

2.1	Constituição do sistema urinário (Duarte, 2014).	7
2.2	Imagens de um rim adulto (University of Utah, 2022).	7
2.3	Processo de filtragem do sangue pelo glomérulo (Universidade Estadual Paulista, 2024).	8
2.4	Imagem de um glomérulo normal.	8
2.5	Profissionais de saúde realizando uma biópsia renal utilizando métodos guiados por imagem como ultrassonografia (National Institute of Diabetes and Digestive and Kidney Diseases (U.S.), 2022).	9
2.6	Imagens de glomérulos sem lesão tingidas com HE e PAS.	10
2.7	Exemplo ilustrativo de sub-conceitos (a) e pequenos disjuntos (b) (Fernandez et al., 2018).	14
2.8	Arquitetura de um modelo Baseline-Ensemble. Na figura, que descreve um <i>ensemble</i> composto por modelos binários, três classificadores-base são utilizados para determinar a saída por meio de uma votação simples.	16
2.9	Curva de Características do Receptor (Fernandez et al., 2018).	19
2.10	Curva Precision-Recall (Fernandez et al., 2018).	20
2.11	Funcionamento de uma Rede Prototípica. As amostras do conjunto de suporte são usadas para a criação dos protótipos com a aplicação da função f_ϕ . Quando uma nova amostra é submetida, esta também é codificada com a função f_ϕ e seu <i>embedding</i> é comparado com cada protótipo, gerando uma pontuação de distância com base na função d . O resultado final da classificação é decidido pela cálculo da menor distância (<i>min</i>).	22
2.12	Os três níveis de desbalanceamento em CIFSL (Ochal et al., 2023).	24
3.1	Distribuição do conjunto utilizado por Raj et al. (2022). O <i>dataset</i> é composto por 3.064 imagens.	27
3.2	Distribuição do conjunto utilizado por Song et al. (2021). O <i>dataset</i> é composto por 3.851 amostras. Fonte: o próprio autor.	28
3.5	Distribuição do conjunto utilizado por Dai et al. (2022). Fonte: o próprio autor.	29
3.6	Distribuição do conjunto utilizado por Haque et al. (2022).	30
3.7	Distribuição do conjunto utilizado por Titoriya e Singh (2022).	32

3.3	Distribuição dos <i>datasets</i> empregados no estudo realizado por Chen et al. (2024b). Os conjuntos CAER-S, RAF-DB e AffectNet utilizados no estudo possuem 65.983, 15.339 e 283.721 amostras respectivamente. Fonte: o próprio autor.	37
3.4	Distribuição do conjunto utilizado por Ilmawati et al. (2024). O conjunto possui um total de 35.887 imagens. Fonte: o próprio autor. . .	38
4.1	Imagens de glomérulos presentes no conjunto de dados. Da esquerda para a direita, são apresentadas amostras tingidas respectivamente com os corantes AZAN, HE, PAMS e PAS. As Figuras 4.1(a) a 4.1(d) apresentam glomérulos sem lesão; 4.1(e) a 4.1(h), glomérulos com <i>amiloidose</i>	41
4.2	Imagens de glomérulos presentes no conjunto de dados. Da esquerda para a direita, são apresentadas amostras tingidas respectivamente com os corantes AZAN, HE, PAMS e PAS. As Figuras 4.2(a) a 4.2(d), <i>esclerose pura sem crescente</i> ; 4.2(e) a 4.2(h), glomérulos com <i>glomerulonefrite crescente</i> ; 4.2(i) a 4.2(l) apresentam glomérulos com <i>nefropatia membranosa</i> ; 4.2(m) a 4.2(p), <i>podocitopatia</i>	42
4.3	Imagens de glomérulos presentes no conjunto de dados. As Figuras 4.3(a) a 4.3(c), <i>esclerose</i> ; 4.3(d) a 4.3(f), <i>hipercelularidade</i> ; 4.3(g) a 4.3(h), <i>hipercelularidade pura sem crescente</i>	43
4.4	Ocupação aproximada desejável de um glomérulo dentro do quadro da imagem.	44
4.5	Distorções encontradas em amostras do conjunto.	45
4.6	Tratamento de distorção de ocupação.	46
4.7	Tratamento para amostras com mais de um glomérulo.	47
4.8	Treinamento episódico balanceado sobre um conjunto desbalanceado. Para cada <i>fold</i> , foram conduzidas até 100 épocas, com 100 episódios por época. A validação final foi realizada sobre dados desbalanceados.	53
5.1	Análise de componentes principais para a classe <i>amiloidose</i>	57
5.2	Distribuição das predições incorretas para <i>amiloidose</i>	59
5.3	Amostras que produziram falsos positivos.	60
5.4	Amostras que produziram falsos negativos.	60
5.5	Análise de componentes principais para a classe <i>normal</i>	61
5.6	Distribuição das predições incorretas para <i>normal</i>	63
5.7	Amostras que produziram falsos positivos.	64
5.8	Amostras que produziram falsos negativos.	64
5.9	Análise de componentes principais para a classe <i>esclerose pura sem crescente</i>	65
5.10	Distribuição das predições incorretas para <i>esclerose pura sem crescente</i>	67
5.11	Amostras que produziram falsos positivos.	68
5.12	Amostras que produziram falsos negativos.	68
5.13	Análise de componentes principais para <i>hipercelularidade</i>	69

5.14	Distribuição das predições incorretas para <i>hiper celularidade</i>	71
5.15	Amostras que produziram falsos positivos.	72
5.16	Amostras que produziram falsos negativos.	72
5.17	Análise de componentes principais para <i>hiper celularidade pura sem crescente</i>	73
5.18	Distribuição das predições incorretas para <i>hiper celularidade pura sem crescente</i>	75
5.19	Amostras que produziram falsos positivos.	75
5.20	Amostras que produziram falsos negativos.	76
5.21	Análise de componentes principais <i>glomerulonefrite crescente</i>	77
5.22	Distribuição das predições incorretas para <i>glomerulonefrite crescente</i>	79
5.23	Amostras que produziram falsos positivos.	79
5.24	Amostras que produziram falsos negativos.	80
5.25	Análise de componentes principais para <i>nefropatia membranosa</i>	81
5.26	Distribuição das predições incorretas para <i>nefropatia membranosa</i>	83
5.27	Amostras que produziram falsos positivos.	83
5.28	Amostras que produziram falsos negativos.	84
5.29	Análise de componentes principais para <i>esclerose</i>	85
5.30	Distribuição das predições incorretas para <i>esclerose</i>	87
5.31	Amostras que produziram falsos positivos.	88
5.32	Amostras que produziram falsos negativos.	88
5.33	Análise de componentes principais para <i>podocitopatia</i>	89
5.34	Distribuição das predições incorretas para <i>podocitopatia</i>	91
5.35	Amostras que produziram falsos positivos.	92
5.36	Amostras que produziram falsos negativos.	92
D.1	Curvas da função de custo para <i>amiloidose</i>	154
D.2	Curvas <i>precision-recall</i> para <i>amiloidose</i>	155
D.3	Matrizes de confusão para <i>amiloidose</i>	155
D.4	Curvas da função de custo para <i>normal</i>	156
D.5	Curvas <i>precision-recall</i> para <i>normal</i>	156
D.6	Matrizes de confusão para <i>normal</i>	157
D.7	Curvas da função de custo para <i>esclerose pura sem crescente</i>	157
D.8	Curvas <i>precision-recall</i> para <i>esclerose pura sem crescente</i>	158
D.9	Matrizes de confusão para <i>esclerose pura sem crescente</i>	158
D.10	Curvas da função de custo para <i>hiper celularidade</i>	159
D.11	Curvas <i>precision-recall</i> para <i>hiper celularidade</i>	159
D.12	Matrizes de confusão para <i>hiper celularidade</i>	160
D.13	Curvas da função de custo para <i>hiper celularidade pura sem crescente</i>	160
D.14	Curvas <i>precision-recall</i> para <i>hiper celularidade pura sem crescente</i>	161
D.15	Matrizes de confusão para <i>hiper celularidade pura sem crescente</i>	161
D.16	Curvas da função de custo para <i>glomerulonefrite crescente</i>	162
D.17	Curvas <i>precision-recall</i> para <i>glomerulonefrite crescente</i>	162
D.18	Matrizes de confusão para <i>glomerulonefrite crescente</i>	163

D.19	Curvas da função de custo para <i>nefropatia membranosa</i>	163
D.20	Curvas <i>precision-recall</i> para <i>nefropatia membranosa</i>	164
D.21	Matrizes de confusão para <i>nefropatia membranosa</i>	164
D.22	Curvas da função de custo para <i>esclerose</i>	165
D.23	Curvas <i>precision-recall</i> para <i>esclerose</i>	165
D.24	Matrizes de confusão para <i>esclerose</i>	166
D.25	Curvas da função de custo para <i>podocitopatia</i>	166
D.26	Curvas <i>precision-recall</i> para <i>podocitopatia</i>	167
D.27	Matrizes de confusão para <i>podocitopatia</i>	167

Lista de Abreviações

Abreviação	Descrição
Acc	<i>Acurácia (Accuracy)</i>
AUC	<i>Area Under the Curve</i>
AUCROC	<i>Area Under the Curve for the Receiver Operating Characteristic</i>
AZAN	<i>Azocarmine–Aniline Blue</i>
BAC	<i>Balanced Accuracy</i>
CAD	<i>Computer Aided Diagnosis</i>
CFAL	<i>Center-Focused Affinity Loss</i>
CHSMOTE	<i>Convex Hull-based Synthetic Minority Oversampling Technique</i>
CI	<i>Class Imbalance</i>
CIFSL	<i>Class Imbalance Few-shot Learning</i>
CNN	<i>Convolutional Neural Networks</i>
CPATH	<i>Computational Pathology</i>
CSL	<i>Cost Sensitive Learning</i>
CR	<i>Congo Red</i>
CV	<i>Computer Vision</i>
DA	<i>Data Augmentation</i>
DL	<i>Deep Learning</i>
FCL	<i>Fully Connected Layer</i>
FE	<i>Feature Extraction</i>
FER	<i>Facial Expression Recognition</i>
FN	<i>False Negative</i>
FP	<i>False Positive</i>
FS	<i>Feature Selection</i>
FSL	<i>Few-shot Learning</i>
GAN	<i>Generative Adversarial Network</i>
HE	<i>Hematoxylin and Eosin</i>
IR	<i>Imbalance Rate</i>
KCN	<i>Kolmogorov–Arnold ConvNeXt Network</i>
LGPD	<i>Lei Geral de Proteção de Dados</i>
MAML	<i>Model Agnostic Meta Learning</i>
MCC	<i>Matthews Correlation Coefficient</i>
ML	<i>Machine Learning</i>

MST	<i>Minimal Spanning Tree</i>
PAMS	<i>Periodic Acid–Methenamine Silver</i>
PAS	<i>Periodic Acid–Schiff</i>
PCA	<i>Principal Component Analysis</i>
PN	<i>Prototypical Network</i>
ROC	<i>Receiver Operating Characteristic</i>
ROS	<i>Random Oversampling</i>
RUS	<i>Random Undersampling</i>
SMOTE	<i>Synthetic Minority Oversampling Technique</i>
SVM	<i>Support Vector Machine</i>
TL	<i>Transfer Learning</i>
TN	<i>True Negative</i>
TP	<i>True Positive</i>
WSI	<i>Whole Slide Imaging</i>

Lista de Símbolos

Símbolo	Descrição
$\approx$	Aproximadamente
d	Função de Cálculo de Distância
$f(\phi)$	Função Codificadora com parâmetros ϕ
ϕ	Parâmetros para a Função Codificadora
ρ	Taxa de Desbalanceamento
$\text{sim}_{\cos}(\mathbf{x}, \mathbf{y})$	Função similaridade do cosseno
$\ \mathbf{x} - \mathbf{y}\ _2$	Função distância euclidiana
N_1	Fração de pontos de borda

Capítulo 1

Introdução

O diagnóstico assistido por computador (*Computer Aided Diagnosis* - CAD), é parte fundamental das aplicações da Patologia Computacional (*Computational Pathology* - CPATH), que é considerada uma abordagem baseada em dados para a Patologia, incluindo, mas não limitada a imagens histológicas e meta-dados de pacientes clínicos, para a análise e extração de conhecimento (Abels et al., 2019). Avanços na digitalização de imagens de lâminas de biópsias têm viabilizado a migração dos métodos de análise tradicional, feitos de forma manual, para técnicas capazes de processar e analisar a imagem de forma automática, com o auxílio de aplicações baseadas em aprendizado de máquina (*Machine Learning* - ML), permitindo que patologistas consigam se concentrar em outras etapas do fluxo de trabalho (Cui e Zhang, 2021), (Calumby et al., 2023). A obtenção de alto desempenho em modelos de ML, especialmente naqueles dedicados ao reconhecimento de padrões em imagens, geralmente demanda grandes conjuntos de dados para treinamento (Abbas, 2023); além disso, os algoritmos empregados por essas ferramentas têm seu desempenho impactado negativamente em um cenário de desbalanceamento de classes, o que ocorre quando pelo menos uma das classes presentes no conjunto é sub-representada, levando a um viés para as classes majoritárias (Wang e Cheng, 2021).

Na Patologia Computacional, a coleta de dados enfrenta limitações práticas significativas, como o elevado custo envolvido na anotação das amostras utilizadas para a construção dos conjuntos de treinamento. Além disso, há importantes questões éticas relacionadas à privacidade, curadoria, tutela e uso dos dados, especialmente quando estes são manipulados por entidades privadas (Abels et al., 2019), (Hosseini et al., 2024). No contexto da legislação brasileira, a Lei nº 13.709 de 2018 — conhecida como Lei Geral de Proteção de Dados Pessoais (LGPD) — classifica os dados de saúde como sensíveis, atribuindo-lhes um regime especial quanto à coleta, tratamento e utilização. Outro fator a ser mencionado é o fato de que os conjuntos de dados médicos tendem a ser, por natureza, desbalanceados (Edward et al., 2023). Esse é o cenário encontrado na classificação de lesões em glomérulos renais, especialmente com respeito a lesões mais raras que, por consequência de sua escassez,

produzem menos amostras histológicas. Lesões pouco frequentes impõem um desafio adicional à construção de ferramentas ML para sua classificação, por não apresentarem amostras suficientes para a generalização dos modelos e ainda por acentuarem o problema de desbalanceamento de dados, quando presentes em conjuntos com amostras de classes mais numerosas (Escobar Díaz Guerrero et al., 2024).

Na literatura acadêmica, o desbalanceamento de classes é um tópico já amplamente discutido, apresentando diferentes técnicas consolidadas para o seu tratamento, que envolvem abordagens orientadas a dados (Megahed et al., 2021), abordagens orientadas algoritmos (Khaleel et al., 2023) e ainda técnicas híbridas (Johnson e Khoshgoftaar, 2019). Nas publicações relacionadas à Patologia Computacional, é possível encontrar diversas aplicações dessas técnicas (Wang e Cheng, 2021), (Mahbub et al., 2024), (Liu et al., 2022).

Outro fator que pode impactar a construção de modelos de ML é o contexto de escassez de dados, especialmente ao se utilizar métodos convencionais, que expostos a esse cenário, tendem a produzir *overfitting*¹ (Parnami e Lee, 2022). Uma técnica encontrada com relativa frequência em trabalhos acadêmicos para tratar cenários de escassez é o aprendizado com poucas amostras (*Few-shot Learning* - FSL) (Fei-Fei et al., 2006), que utiliza conhecimento prévio e técnicas específicas, como *redes prototípicas* e *meta-aprendizado* (*meta learning*), para permitir a um modelo o aprendizado de novas classes a partir de poucos dados. Em alguns estudos, a aplicação de FSL também é citada como uma possível abordagem para cenários de desbalanceamento, podendo até mesmo ser combinada com estratégias comuns (Billion Polak et al., 2024), (Ochal et al., 2023).

Além do desbalanceamento de classes e da escassez de dados, dois fatores que já impõem desafios significativos ao treinamento de modelos de aprendizado de máquina em tarefas de classificação, outro aspecto igualmente relevante diz respeito à qualidade e à complexidade do conjunto de dados. Fernandez et al. (2018) argumentam que o desbalanceamento, por si só, não é o único fator determinante para o desempenho dos classificadores, destacando a influência de características intrínsecas ao próprio conjunto de dados empregado em cada tarefa.

No estudo desenvolvido neste trabalho, foram explorados diferentes cenários de desbalanceamento e complexidade de dados no contexto da classificação automática de lesões glomerulares. Para isso, utilizou-se um conjunto composto por 11.674 amostras de imagens histológicas de glomérulos renais, com o objetivo de adotar uma abordagem prática e alinhada a contextos reais e heterogêneos, tanto em termos de desbalanceamento entre classes quanto de complexidade intrínseca dos dados.

O conjunto utilizado contempla amostras de oito diferentes tipos de lesões, além de glomérulos sem lesão ou *normais*, totalizando nove classes. As amostras presentes no conjunto foram tingidas com quatro diferentes corantes histológicos. Cada

¹Quando um modelo se adapta aos dados aprendidos, sem contudo extrair padrões que permitam a sua generalização, apresentando baixo desempenho ao ser utilizado sobre novos dados (Goodfellow et al., 2016).

classe apresenta uma taxa distinta de desbalanceamento, chegando a aproximadamente 1:52 para a classe menos representada. Esse nível de desbalanceamento é expressivo, como será evidenciado na comparação com outros estudos discutidos no Capítulo 3. Salienta-se que, a partir do conjunto original, foram construídas diferentes configurações para as tarefas de aprendizado para cada classe.

O estudo aqui apresentado empregou diferentes técnicas e procedimentos para tratar o desbalanceamento presente em cada configuração do conjunto utilizado para o treinamento dos modelos, além de estratégias para reduzir a potencial complexidade existente em cada uma dessas configurações.

Como será melhor descrito no Capítulo 4, para tratar o desbalanceamento de classes foram adotadas as abordagens mais comuns presentes na literatura acadêmica, utilizando-se redes neurais convolucionais (*Convolutional Neural Networks* - CNN). Um comparativo entre o uso de técnicas convencionais e FSL também foi realizado, com o objetivo de identificar a melhor técnica ou combinação de técnicas que mais se adequaria a cada cenário.

Apesar de FSL ter sido originalmente projetada para contextos de extrema escassez de dados, sua aplicação neste trabalho se justifica pela baixa representatividade de algumas classes frente ao conjunto estudado, o que torna a construção de modelos convencionais desafiadora. Os resultados apresentados no Capítulo 5 ressaltam a aplicabilidade de FSL quando selecionado um número limitado de amostras para o treinamento de modelos baseados em redes prototípicas.

No tocante ao tratamento da complexidade dos conjuntos, foi realizada uma segmentação por corante, com o intuito de se reduzir o potencial impacto negativo trazido por amostras de corantes distintos. Aspectos concernentes à qualidade das amostras também foram tratados, por meio de um processo de curadoria manual, que buscou identificar e reduzir potenciais ruídos presentes no conjunto.

Como será mostrado no Capítulo 5, os resultados apresentaram taxas de F1-score acima de 90% para classes com desbalanceamento da ordem de $\approx 1:30$, chegando a $\approx 96\%$ para amostras sem lesão. Entretanto, classes com relativa maior representação apresentaram modelos com desempenho inferior, evidenciando que outros fatores, além da taxa de desbalanceamento tiveram influência sobre o desempenho.

Por se tratar de um estudo comparativo entre diferentes técnicas, abrangendo desde o tratamento dos dados até a aplicação de distintos modelos de aprendizado, foi dada atenção especial à significância estatística na comparação dos resultados obtidos. Conforme detalhado no Capítulo 4, avaliou-se a relevância estatística das diferenças de desempenho observadas entre as configurações adotadas no treinamento dos modelos, a fim de garantir que as conclusões fossem sustentadas por evidências quantitativas.

1.1 Objetivos

Esta seção apresenta os objetivos deste estudo, partindo do objetivo principal e passando pelos objetivos específicos identificados.

1.1.1 Objetivo Principal

O objetivo principal deste trabalho foi desenvolver, otimizar e avaliar sistematicamente classificadores automáticos para a classificação de lesões em glomérulos renais, enfrentando o cenário imposto pelo severo desbalanceamento de classes e pela complexidade intrínseca dos dados histopatológicos. Para tanto, buscou-se comparar e integrar abordagens avançadas de *Few-Shot Learning* com técnicas convencionais de *Machine Learning*, buscando um desempenho elevado e robusto, especialmente na classificação de lesões raras, no contexto da Nefropatologia Computacional.

1.1.2 Objetivos Específicos

Os objetivos específicos deste estudo estão definidos como a seguir:

- Analisar comparativamente a eficácia de abordagens baseadas em *Few-Shot Learning* e em Redes Neurais Convolucionais convencionais para o tratamento do desbalanceamento severo de classes, identificando os cenários mais adequados para a aplicação de cada técnica ou suas combinações;
- Investigar o impacto da qualidade dos dados, por meio de um processo de curadoria manual para verificação detalhada de como alterações na apresentação dos glomérulos nas imagens influenciam no desempenho e na capacidade de generalização dos classificadores;
- Avaliar a influência da segmentação de dados por tipo de corante histológico (multi-corante versus mono-corante) no desempenho dos modelos.

1.2 Contribuições

Apesar dos avanços na classificação de lesões glomerulares, a literatura carece de uma abordagem que integre um processo detalhado de curadoria de dados e a análise do impacto de corantes específicos no treinamento de modelos de ML. Neste sentido, este estudo inova ao propor e avaliar sistematicamente essa combinação para otimizar a classificação de lesões glomerulares.

As principais contribuições deste trabalho são:

- Uma avaliação comparativa de classificadores automáticos para lesões glomerulares, com objetivo de alcançar desempenho superior a 85% para a métrica F1-score em cenários de desbalanceamento de $\approx 1:30$, utilizando uma combinação de FSL e CSL.

- Um estudo detalhado da aplicação de métodos tradicionais de ML em um cenário de desbalanceamento acentuado dentro da Patologia Computacional, mais especificamente na nefropatologia, com uma análise comparativa dos resultados e observações sobre os desafios e oportunidades encontrados
- Uma análise comparativa entre as técnicas convencionais e o uso de FSL, explorando alternativas para a construção de modelos baseados em redes prototípicas em cenários de desbalanceamento (*Class Imbalance Few-shot Learning* - CIFSL), identificando as condições em que cada abordagem se mostra mais eficaz.
- Um estudo prático sobre a influência da qualidade das amostras para a construção dos classificadores, através de um processo de curadoria do conjunto original e comparação de resultados, demonstrando o impacto da remoção de ruídos e artefatos na capacidade de generalização dos modelos.

1.3 Organização do Trabalho

Este texto apresenta, no Capítulo 2, uma fundamentação dos principais tópicos aqui abordados. São introduzidos os conceitos sobre *Machine Learning* e Patologia Computacional e são apresentados os tópicos sobre Desbalanceamento de Classes e Escassez de Dados. O Capítulo 3 apresenta os trabalhos relacionados e seus resultados para fins comparativos. O Capítulo 4 apresenta a metodologia adotada neste trabalho, a construção dos conjuntos de dados e os procedimentos para a realização dos experimentos. O Capítulo 5 apresenta os resultados alcançados e discute os achados e observações. Finalmente o Capítulo 6 conclui este estudo e apresenta propostas para trabalhos futuros.

Capítulo 2

Fundamentação Teórica

Esta seção apresenta os principais conceitos que fundamentam este trabalho e são necessários ao seu entendimento, iniciando por fundamentos da Nefropatologia, passando por fundamentos da Computação, Patologia Computacional e Aprendizado de Máquina.

2.1 Conceitos de Nefropatologia

2.1.1 Rins e Glomérulo

O sistema urinário é constituído por um par de rins, de ureteres, uma bexiga e uma uretra (Oliveira e Neto, 2015), como representado na Figura 2.1. Os rins realizam a maior parte das funções de excreção, filtrando o sangue e recolhendo deste os resíduos metabólicos de todas as células do nosso corpo (Dio, 1999), (Gardner et al., 1988), (Moore e Dalley, 2007). A urina produzida pelos rins, é o veículo por meio do qual são eliminados a água, sais minerais, íons, resíduos metabólicos, enfim, substâncias que em excesso causam um desequilíbrio fisiológico em nosso organismo (Moraes e Colicigno, 2007).

Os rins estão localizados no espaço retroperitoneal da parede abdominal posterior, cada um do lado da coluna vertebral, na altura da 12^a vértebra torácica à terceira vértebra lombar (Montanari, 2016). Cada rim pesa cerca de 150g e mede 10-12cm de comprimento, 4-6,5cm de largura e 2-3cm de espessura e seu formato assemelha-se ao de um carço de feijão. A Figura 2.2 exibe a imagem de um rim normal de uma pessoa adulta, à direita, uma seção transversal revelando as estruturas internas do órgão.

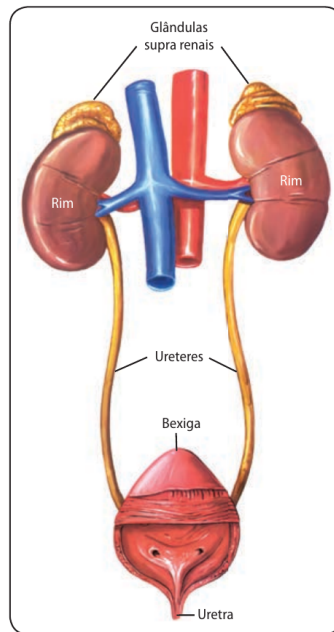


Figura 2.1: Constituição do sistema urinário (Duarte, 2014).



(a) Rim normal de uma pessoa adulta



(b) Corte transversal de um rim

Figura 2.2: Imagens de um rim adulto (University of Utah, 2022).

O glomérulo é uma estrutura vascular componente do néfron, que é a unidade morfofuncional básica do rim. Possuem capilares fenestrados e uma espessa membrana através da qual o sangue é filtrado, separando aquilo será reabsorvido e o que será eliminado. Esse processo de filtração é a primeira etapa da produção da urina (Mo-
raes e Colicigno, 2007), (Montanari, 2016), (Universidade Estadual Paulista, 2024). A Figura 2.3 ilustra o funcionamento de um néfron, com o processo de filtragem do sangue pela estrutura glomerular. A Figura 2.4 exibe um glomérulo normal.

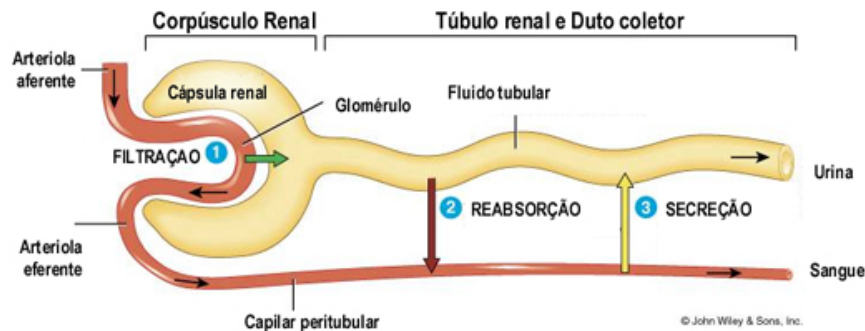


Figura 2.3: Processo de filtragem do sangue pelo glomérulo (Universidade Estadual Paulista, 2024).

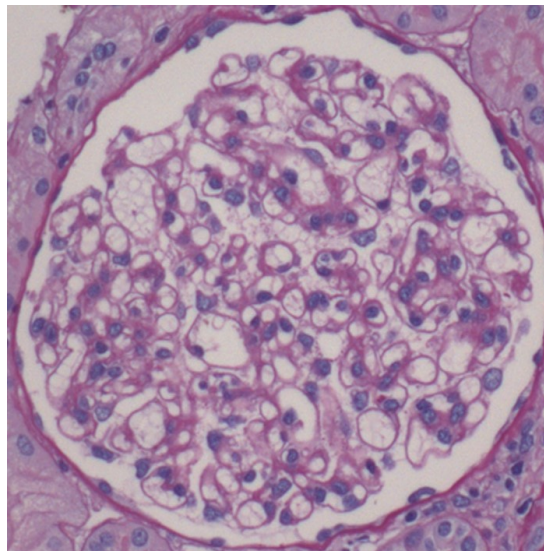


Figura 2.4: Imagem de um glomérulo normal.

2.1.2 Biópsia

Para o diagnóstico de lesões glomerulares, é necessária a realização de uma *biópsia* (Monteiro e Diz, 2015). Uma biópsia é um procedimento médico relativamente simples que envolve a retirada de uma pequena amostra de tecido orgânico para

posterior análise histológica (Whittier e Korbet, 2004). A Figura 2.5 ilustra uma biópsia renal guiada por imagem.

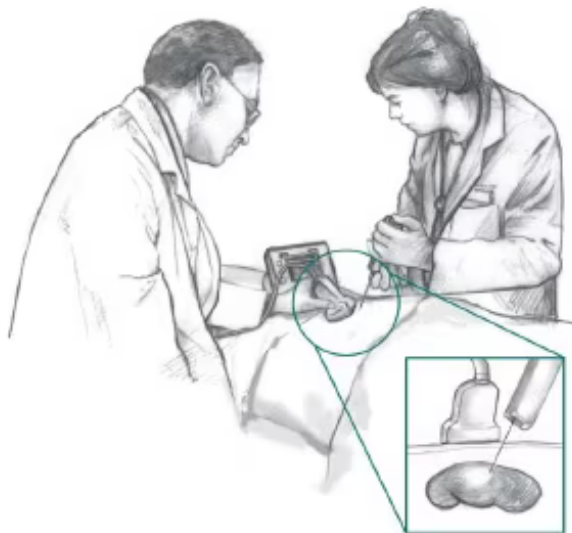
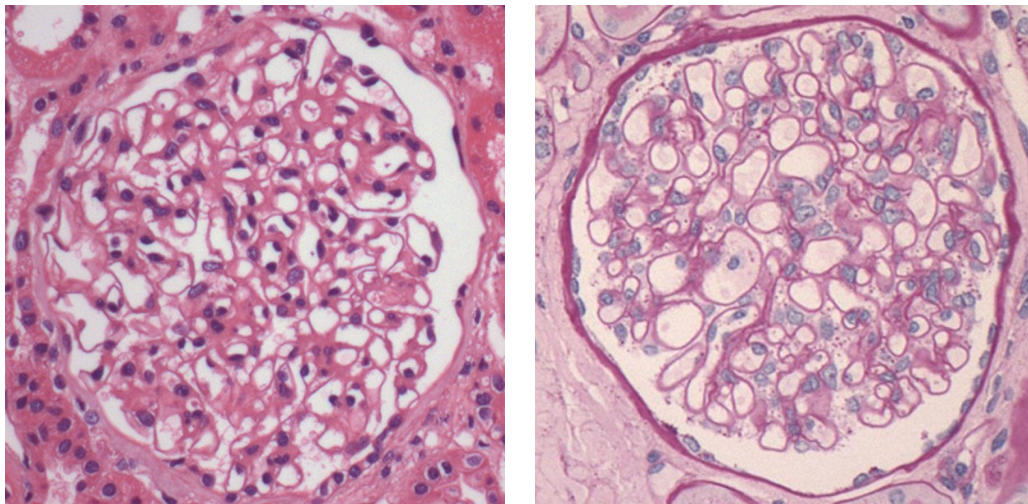


Figura 2.5: Profissionais de saúde realizando uma biópsia renal utilizando métodos guiados por imagem como ultrassonografia (National Institute of Diabetes and Digestive and Kidney Diseases (U.S.), 2022).

Antes da análise do material coletado na biópsia em um microscópio, é necessário que os tecidos orgânicos passem por um processo de coloração, permitindo a visualização das suas estruturas e mudanças em suas características (Gurina e Simms, 2023). Nos experimentos conduzidos neste estudo, foi dada ênfase maior no uso de amostras tingidas com os corantes Hematoxilina e Eosina (HE) e Ácido Periódico de Schiff (PAS), por serem estes os mais utilizados em análises histológicas (Timm, 2005), sendo também os mais presentes no conjunto original aqui utilizado. A Figura 2.6 exhibe dois glomérulos saudáveis, obtidos em imagens de biópsias tingidas com os corantes HE e PAS.



(a) Imagem de um glomérulo sem lesão tingida com HE (b) Imagem de um glomérulo sem lesão tingida com PAS

Figura 2.6: Imagens de glomérulos sem lesão tingidas com HE e PAS.

2.1.3 Lesões Glomerulares

Este estudo é realizado sobre um conjunto de dados contendo 8 tipos de lesões glomerulares, como descrito a seguir:

- **Amiloidose:** Refere-se a um termo aplicado a um grupo heterogêneo de doenças raras, caracterizado pelo depósito extracelular de proteínas amiloides, causando disfunção de órgãos-alvo e uma ampla gama de sintomas clínicos, tais como fadiga, perda de peso insuficiência cardíaca e gastrointestinal (Barbosa et al., 2015), (Monteiro e Diz, 2015).
- **Esclerose (Glomeruloesclerose) e Esclerose Pura:** Lesão caracterizada pela formação de cicatrizes pela interação de células glomerulares e material extracelular. O termo *esclerose pura* refere-se ao fato de a lesão ocorrer isoladamente, não sendo consequência direta de alguma outra condição patológica (Goldenberg e Carvalho, 2012).
- **Hipercelularidade e Hipercelularidade Pura:** Lesão caracterizada pelo aumento no número de células nas estruturas glomerulares. O termo *hipercelularidade pura* também sugere que a doença aparece sem a presença de outras patologias (Chagas et al., 2020).
- **Glomerulonefrite crescente:** É uma doença imunomediada crônica que causa inflamação e lesão glomerular grave. Frequentemente leva à insuficiência renal irreversível (Al Mushafi et al., 2021).
- **Nefropatia membranosa:** Lesão caracterizada por um espessamento aparente das paredes capilares glomerulares, que resulta da formação de complexos imunes no aspecto externo da membrana basal (Ronco et al., 2021).

- **Podocitopatia:** Doença renal na qual a lesão direta ou indireta dos podócitos causa proteinúria ou síndrome nefrótica (Kopp et al., 2020). Podócitos são células altamente especializadas do glomérulo renal que envolvem capilares e são vizinhas das chamadas células de Bowman (Reiser e Altintas, 2016).

2.2 Conceitos de Computação

2.2.1 Aprendizado de Máquina

O Aprendizado de Máquina (do inglês, *Machine Learning* - ML) pode ser definido como um conjunto de métodos capazes de detectar padrões em dados, sendo uma área de estudo muito relacionada à Estatística (Murphy, 2012). A construção ou treinamento de modelos de ML geralmente é orientada a estimar funções complexas, que possam representar o comportamento dos dados de entrada (Goodfellow et al., 2016).

As abordagens para ML são geralmente divididas em dois grandes grupos, a saber, os métodos de aprendizado supervisionado ou preditivos, nos quais se procura mapear entradas para saídas, dado um conjunto de dados rotulados¹, e os métodos não-supervisionados ou descritivos, que realizam o treinamento sobre dados sem rótulos com a finalidade de descobrir padrões de interesse sobre esses dados. Dentre os métodos supervisionados, encontram-se as tarefas de classificação, cujo objetivo é mapear uma classe de saída para uma determinada amostra de entrada, sendo uma das formas de ML mais largamente utilizadas na prática (Murphy, 2012).

Na Visão Computacional, a classificação de imagens constitui uma tarefa importante com inúmeras aplicações. Os métodos mais avançados fazem uso de CNNs para a extração de características e aprendizado dos modelos (Toennies, 2024), sendo empregados em cenários diversos como tarefas de diagnóstico (Wang e Cheng, 2021), sensoriamento remoto (Wen Xie e Sun, 2024), realidade aumentada e realidade virtual (Cortes et al., 2024).

2.2.2 Visão Computacional

A Visão Computacional (*Computer Vision* - CV) é um campo interdisciplinar que estuda como reproduzir em um computador a habilidade enxergar, buscando entender como a percepção visual funciona e construindo sistemas capazes de interpretar o mundo ao redor com uso de imagens (Torralba et al., 2024). É uma das áreas de pesquisa mais ativas dentro do Aprendizado de Máquina, compreendendo uma vasta gama de aplicações que vão desde reproduzir a visão humana, até a criação de categorias de habilidades visuais totalmente novas (Goodfellow et al., 2016). Na Patologia, a Visão Computacional tem um papel importante por contribuir para o

¹Com a existência dos rótulos é possível aferir o desempenho dos modelos, já que a saída esperada é conhecida

aumento da produtividade de patologistas, servindo para a construção de ferramentas assistivas ao diagnóstico (Cui e Zhang, 2021).

2.2.3 Redes Neurais Convolucionais

Redes Neurais Convolucionais (*Convolutional Neural Networks* - CNN) são um tipo especializado de rede neural projetado para processar dados com topologia matricial, na qual os dados de entrada são representados em uma grade de células valoradas, com acontece com os *pixels* em uma imagem, sendo amplamente utilizadas na extração de padrões em visuais (Goodfellow et al., 2016). O termo *convolucional* faz referência a uma operação matemática fundamentada na *Álgebra Linear* (Wu, 2017).

Em uma rede convolucional, os dados de entrada são tipicamente submetidos às operações de convolução e *pooling* para extração de características, sendo entregues à camada totalmente conectada (*Fully Connected Layer* - FCL), que mapeia as características extraídas para a saída final da rede (Yamashita et al., 2018). Essa combinação de operações torna as CNNs capazes de extrair características hierárquicas em imagens, justificando sua ampla aplicação na classificação de imagens histológicas.

2.2.4 Desbalanceamento de Classes

Para a construção de modelos de classificação, a maioria dos algoritmos existentes frequentemente assume que o treinamento é realizado sobre dados cuja representatividade de cada classe é relativamente homogênea, tendo o seu desempenho impactado negativamente quando uma ou mais classes são pouco representadas ou até mesmo raras dentro do conjunto, em um cenário conhecido como *desbalanceamento de classes* (*class imbalance* - CI) (Fernandez et al., 2018). Cenários de desbalanceamento de classes, em regra, contribuem para a degradação de modelos de ML que, por criarem um viés para a classe majoritária (aquela de maior ocorrência), apresentam um desempenho inferior no aprendizado da classe minoritária (Megahed et al., 2021). O problema de desbalanceamento pode se apresentar em diversos níveis, que estão intrinsecamente relacionados com a natureza do conjunto em análise (Sakri e Basheer, 2023).

No âmbito da Patologia computacional, os problemas relacionados ao desbalanceamento de classes também são percebidos em aplicações para auxílio ao diagnóstico por imagem (Walsh e Tardy, 2023), (Mahbub et al., 2024). Wang e Cheng afirmam que obter balanceamento em conjuntos de dados médicos pode ser desafiador devido à heterogeneidade presente em diversos aspectos do conjunto, alta variabilidade dos dados, ou mesmo questões relacionadas à quantidade de pacientes (Wang e Cheng, 2021). Esses aspectos dos conjuntos médicos são também percebidos no conjunto utilizado nesta pesquisa, em que existem classes com representação muito inferior às demais, o que será melhor detalhado na descrição do *dataset*, no capítulo 4.

Nos trabalhos referenciados neste estudo, não foi encontrado um consenso no tocante à classificação de conjuntos de dados com base em sua taxa de desbalanceamento (*imbalance ratio* - IR)². Walsh e Tardy (2023), por exemplo, consideraram alta uma IR de 1:19, enquanto Sakri e Basheer (2023) consideraram como alto apenas um desbalanceamento acima de 1:100, Mahbub et al. (2024) nomearam como *pesado*, um desbalanceamento de $\approx 1:3$, já Nourivandi et al. (2024), que trabalharam em um problema multiclasse, apontaram um alto desbalanceamento com duas classes compondo 70% dos dados. A Equação 2.1 apresenta o cálculo para IR, onde o numerador é a quantidade de amostras da classe majoritária e o denominador é a quantidade de amostras da classe minoritária (Ochal et al., 2023).

$$IR = \frac{\text{numero_de_amostras_da_classe_majoritaria}}{\text{numero_de_amostras_da_classe_minoritaria}} \quad (2.1)$$

2.2.5 Complexidade dos Dados

Além do Desbalanceamento de Classes, outro fator que deve ser considerado em uma tarefa de classificação é a complexidade dos dados. Segundo Fernandez et al. (2018), o desbalanceamento isoladamente não necessariamente é um fator determinante para a dificuldade de uma tarefa de classificação. Os autores afirmam que podem existir características intrínsecas nos dados com potencial impacto negativo no desempenho dos modelos. Duas das principais dessas características são os sub-conceitos (*sub-concepts*) e os pequenos disjuntos (*small disjuncts*). Sub-conceitos ocorrem quando uma classe não pertence a uma região homogênea do espaço de características, pela existência de particularidades que afetam sub-grupos dentro de uma mesma classe, permitindo que esta seja separada ou classificada de diferentes formas. Pequenos disjuntos referem-se à presença escassa de amostras de uma classe em certas regiões do espaço de características. A Figura 2.7, extraída de Fernandez et al. (2018) traz uma ilustração desses problemas, com uma distribuição sem a presença de sub-conceitos à esquerda (a). O pequeno grupo, localizado no canto inferior direito da Figura 2.7 (b) representa, além de um sub-conceito, também um pequeno disjunto.

²Este texto utiliza a notação 1:IR para referenciar o grau ou taxa de desbalanceamento.

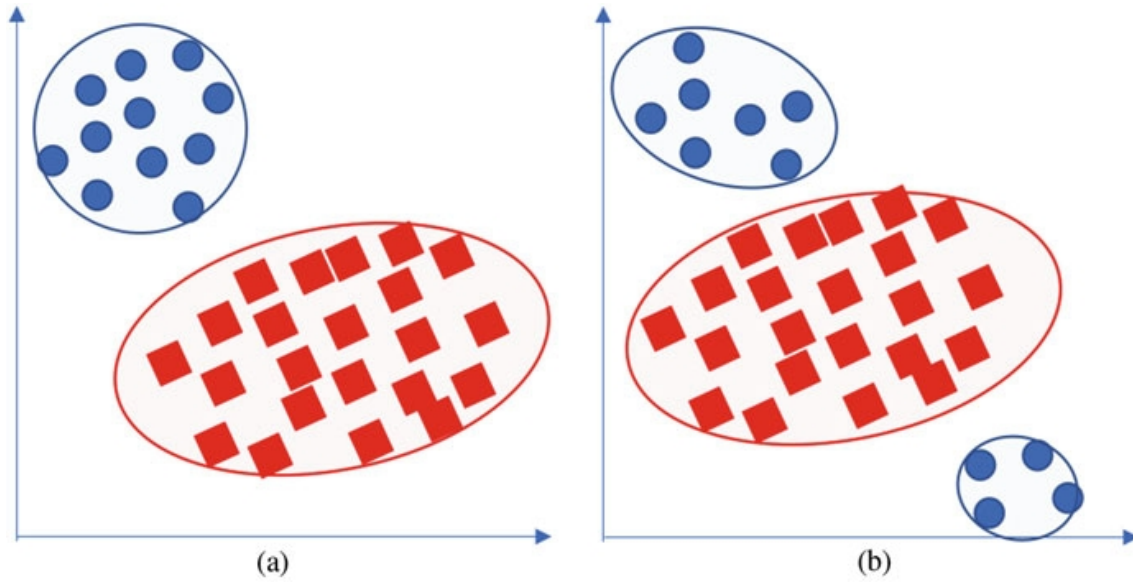


Figura 2.7: Exemplo ilustrativo de sub-conceitos (a) e pequenos disjuntos (b) (Fernandez et al., 2018).

Considerando-se a relevância da complexidade dos dados para tarefas de *machine learning* supervisionadas, torna-se importante definir meios para avaliar os conjuntos de dados utilizados nessas tarefas. Dentre o ferramental existente para esse tipo de análise, encontram-se métodos que fornecem informação visual e também aqueles que permitem o cálculo de alguma métrica de complexidade do conjunto estudado. Alguns desses métodos são apresentados a seguir:

- **Análise de componentes principais (*principal component analysis* - PCA):** Trata-se de um técnica baseada na redução da dimensionalidade dos *pontos de dados*, permitindo a extração dos padrões dominantes existentes nos dados e a sua representação de forma gráfica (Wold et al., 1987).
- **Pontuação de Silhueta (*Silhouette Score*):** Métrica utilizada para avaliar a qualidade de agrupamentos, baseada na média entre duas medidas para cada amostra: (i) a distância média intra-agrupamento (*intra-cluster*), que quantifica o quão próxima a amostra está das demais no mesmo grupo; e (ii) a menor distância média em relação aos agrupamentos vizinhos, pertencentes a classes diferentes. Valores mais próximos de 1 indicam melhor separação entre os grupos, valores próximos de 0 sugerem sobreposição entre os agrupamentos e valores próximos de -1 indicam que as amostras foram atribuídas ao agrupamento incorreto (Starczewski e Krzyżak, 2015; Pedregosa et al., 2011). A fórmula para o cálculo do índice é apresentada na Equação 2.2, na qual, a é a média da distância *intra-cluster* e b é a média da distância para o *cluster* vizinho.

$$Silhouette_score = \frac{(b - a)}{\max(a, b)} \quad (2.2)$$

- **Fração de pontos de borda (*fraction of borderline points* - N1):** A partir do agrupamento dos pontos de dados, produz uma árvore geradora mínima (*minimal spanning tree* - MST), a partir da qual, calcula-se o percentual de arestas incidentes oriundas da classe oposta, permitindo medir o tamanho e a complexidade da fronteira de decisão (Lorena et al., 2019). Valores próximos de 0 indicam separação clara entre as classes, enquanto valores próximos de 1 sugerem áreas de sobreposição. A fórmula para o cálculo de N1 é apresentada na Equação 2.3, em que n é o total de arestas, (x_i, x_j) representa cada aresta pertencente à MST e $y_i \neq y_j$ indica que as classes conectadas por essas arestas devem ser diferentes.

$$N1 = \frac{1}{n} \sum_{i=1}^n I((x_i, x_j) \in MST \wedge y_i \neq y_j) \quad (2.3)$$

- **Medidas de linearidade:** Oferecem meios quantitativos de avaliar se as classes do conjunto são linearmente separáveis, ou seja, se podem ser separadas por um hiperplano (Lorena et al., 2019). Orriols-Puig et al. (2010) propuseram a obtenção dessa métrica através da observação do desempenho obtido por uma máquina de vetor de suporte (*Support Vector Machine* - SVM).

2.2.6 Abordagens mais comuns para tratar o Desbalanceamento de Classes

O desbalanceamento de classes é um tópico largamente explorado na literatura acadêmica, apresentando diversas técnicas para o seu enfrentamento (Fernandez et al., 2018). Dentre as mais comuns e amplamente utilizadas, encontramos as abordagens baseadas em dados, as abordagens baseadas em algoritmos e ainda os métodos híbridos (Johnson e Khoshgoftaar, 2019).

Abordagens baseadas em dados aplicam técnicas de reamostragem para tornar o conjunto de dados balanceado ou deixá-lo com uma distribuição que seja mais conveniente para o treinamento, sendo basicamente técnicas aplicadas na etapa de pré-processamento dos dados. Duas das técnicas de reamostragem mais comuns são a sub-amostragem aleatória (*Random-undersampling* - RUS), a superamostragem aleatória (*Random-oversampling* - ROS). RUS realiza a remoção de amostras da classe majoritária, até que esta apresente a mesma distribuição encontrada na classe minoritária; ROS executa a replicação de amostras da classe minoritária, igualando-a em número à classe majoritária. Técnicas de reamostragem híbrida combinam o uso de ROS e RUS para equilibrar o conjunto de dados. Embora muito utilizadas, essas técnicas também podem apresentar desvantagens, uma vez que, estratégias de

superamostragem favorecem a construção de modelos supertreinados (*overfitting*), enquanto métodos de sub-amostragem podem descartar amostras com características mais relevantes para a generalização dos modelos (Fernandez et al., 2018).

Além das técnicas de reamostragem, outro método baseado em dados bastante utilizada é o *aumento de dados* (*Data Augmentation* - DA) que, no contexto da visão computacional, consiste geralmente em alguma transformação aplicada a cópias das imagens existentes, incluindo, mas não limitada a corte, inversão horizontal ou vertical e rotação (Xu et al., 2023).

Os métodos baseados em algoritmos utilizam estratégias para melhorar modelos de classificação de maneira a adaptá-los aos dados desbalanceados (Dai et al., 2023). O uso efetivo dessas técnicas requer um conhecimento avançado sobre detalhes do algoritmo selecionado, para identificar quais de seus mecanismos são responsáveis por criar um viés para a classe majoritária (Fernandez et al., 2018). Um grupo específico de abordagens algorítmicas são as técnicas baseadas em aprendizado sensível ao custo, que alteram a forma como os erros de classificação são calculados, atribuindo maior importância para a classe minoritária (Fernandez et al., 2018).

Outras abordagens para tratar o desbalanceamento de classes incluem os métodos híbridos, que combinam as técnicas baseadas em dados e aquelas baseadas em algoritmos. Também são referenciados na literatura os *modelos baseados em conjunto* (*ensemble-based*), como o descrito na Figura 2.8, que se valem da combinação dos resultados de dois ou mais modelos, chamados de classificadores-base, e também são muito empregados na prática, já que o seu desempenho costuma ser superior ao dos modelos utilizados isoladamente (Fernandez et al., 2018). São citadas ainda outras técnicas relacionadas à *redução de dimensionalidade*, como a *seleção* ou *extração de características* (*Feature Selection* - FS, *Feature Extraction* - FE), que intentam alterar ou transformar o espaço de características das amostras para um que seja mais representativo e passível de ser aprendido pelos modelos de ML, procurando dar mais ênfase às características que identifiquem melhor a classe minoritária (Fernandez et al., 2018).

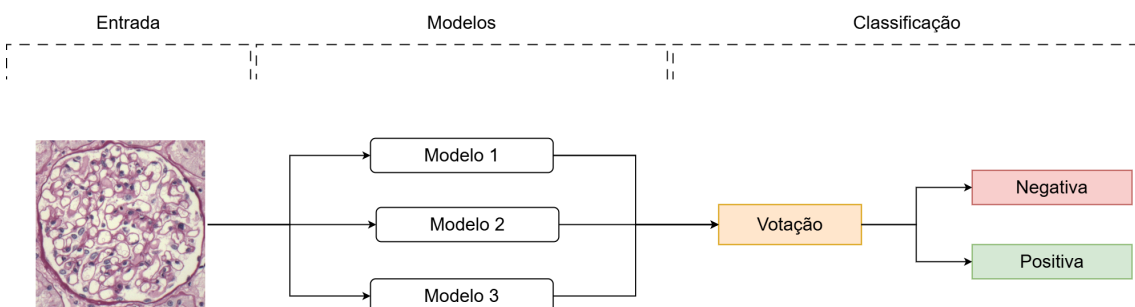


Figura 2.8: Arquitetura de um modelo Baseline-Ensemble. Na figura, que descreve um *ensemble* composto por modelos binários, três classificadores-base são utilizados para determinar a saída por meio de uma votação simples.

2.2.7 Métricas de Desempenho em Cenários de Desbalanceamento

A avaliação de modelos em tarefas de classificação geralmente é feita com valores oriundos do que é chamado de *matriz de confusão*, que, no contexto de uma tarefa binária, é uma tabela 2×2 , na qual, as linhas representam os rótulos ou classes reais (*ground truth*) e as colunas representam o que foi previsto por um modelo. A Tabela 2.1 apresenta um exemplo de matriz de confusão para um problema de classificação binário, em que TP representa a quantidade de amostras positivas classificadas como positivas por um modelo, ou verdadeiros positivos (*True Positives*), TN representa o número de amostras negativas preditas como negativas, ou, os verdadeiros negativos (*True Negatives*), FP representa as amostras negativas que foram classificadas como positivas, os falsos positivos (*False Positives*) e FN conta as amostras positivas que foram previstas como negativas, os falsos negativos (*False Negatives*). As métricas aqui apresentadas, consideram apenas tarefas de classificação binária, salvo se explicitamente disposto de outra maneira.

Tabela 2.1: Matriz de Confusão para um problema de classificação binário.

		Classe Predita	
		Positiva	Negativa
Classe Real	Positiva	TP	FN
	Negativa	FP	TN

A *acurácia* (*accuracy* - Acc), cuja fórmula é dada na Equação 2.4, e a *taxa de erro* (Equação 2.5) são as duas métricas mais comumente utilizadas em tarefas de classificação (Johnson e Khoshgoftaar, 2019). Entretanto, o uso da acurácia é desencorajado no contexto de desbalanceamento pois, uma representatividade exacerbada de uma das classes cria um viés, por levar o modelo produzido a considerar uma possível grande quantidade de acertos para a classe majoritária, ignorando os erros de classificação para as classes sub-representadas (Megahed et al., 2021).

$$Acc = \frac{TP + TN}{TP + FN + TN + FP} \quad (2.4)$$

$$Error_rate = 1 - Acc \quad (2.5)$$

Assim, no tratamento de tarefas com a presença de desbalanceamento de classes, outras métricas são necessárias para corretamente aferir o desempenho dos modelos (Erickson e Kitamura, 2021), (Johnson e Khoshgoftaar, 2019), (Fernandez et al., 2018):

Recall: *revocação* (*recall*), também conhecida como *sensibilidade*, considera-se como a taxa de verdadeiros positivos (Equação 2.6).

$$Recall = \frac{TP}{TP + FN} \quad (2.6)$$

Specificity: *especificidade* ou *seletividade*, representa a taxa de verdadeiros negativos (Eq. 2.7).

$$Specificity = \frac{TN}{TN + FP} \quad (2.7)$$

Precision: *precisão*, representa a fração das previsões positivas corretas (Equação 2.8)

$$Precision = \frac{TP}{TP + FP} \quad (2.8)$$

MCC: *Matthews Correlation Coefficient* é uma medida oriunda da *Bioinformática*, onde o desbalanceamento ocorre com muita frequência, é uma adaptação do Coeficiente de Correlação de Pearson (Fernandez et al., 2018) (Eq. 2.9). A métrica apresenta valores compreendidos no intervalo $[-1, 1]$, no qual, -1 significa que as previsões estão todas erradas, valores próximos a zero sugerem previsões aleatórias e 1 representa que todas as previsões estão corretas (Chicco et al., 2021).

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \quad (2.9)$$

F1-score: também conhecida como *F-measure*, é uma métrica focada na análise da classe positiva, mas especificamente, na relação de equilíbrio entre precisão e sensibilidade, sendo uma média harmônica entre essas duas métricas (Equação 2.10). Em tarefas multiclasse, são utilizadas variantes de F1-score com base em médias calculadas sobre a métrica para as classes do conjunto (Pedregosa et al., 2011):

- *Micro F1-score* calcula a métrica globalmente, contando o total de verdadeiros positivos, falsos negativos e falsos positivos. Essa métrica pode introduzir um viés para as classes majoritárias, uma vez que a média é calculada para todo o sistema, sendo mais indicada quando o desempenho é consistentemente similar entre as classes do conjunto.
- *Macro F1-score* calcula a métrica para cada classe, obtendo uma média simples. É uma medida útil para se avaliar o desempenho sobre cada classe, evitando introduzir um viés para as classes majoritárias, uma vez que cada classe é tratada de forma igual e independente.

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \quad (2.10)$$

Os experimentos realizados neste trabalho utilizaram F1-Score como principal métrica de avaliação dos modelos produzidos.

AUCROC: *área sob a curva de características operacionais do receptor* (*area under the receiver operating characteristics*) é um método gráfico utilizado para ilustrar a compensação existente entre a sensibilidade e a *taxa de falsos positivos* ($1 - \text{specificity}$) (Erickson e Kitamura, 2021). A curva é traçada considerando-se diferentes valores para um *limiar de classificador* (*threshold*), utilizado para estabelecer um limite a partir do qual uma amostra é considerada como pertencente à classe positiva ou negativa. O cálculo da *área sob a curva ROC* é uma métrica bastante comum para CI em tarefas binárias, entretanto, seu uso não é recomendado em cenários de extremo desbalanceamento, em que a classe minoritária é considerada rara, por conta de seu comportamento tornar-se menos sensível aos falsos positivos à medida que quantidade de amostras negativas cresce, apresentando resultados muito otimistas (Johnson e Khoshgoftaar, 2019). A Figura 2.9, extraída de Erickson e Kitamura (2021), exibe uma curva ROC. Quanto mais próximos os valores estiverem do canto superior esquerdo, melhor o desempenho geral do modelo. É válido ressaltar que cada ponto na curva é resultado da escolha de um limiar para o classificador.

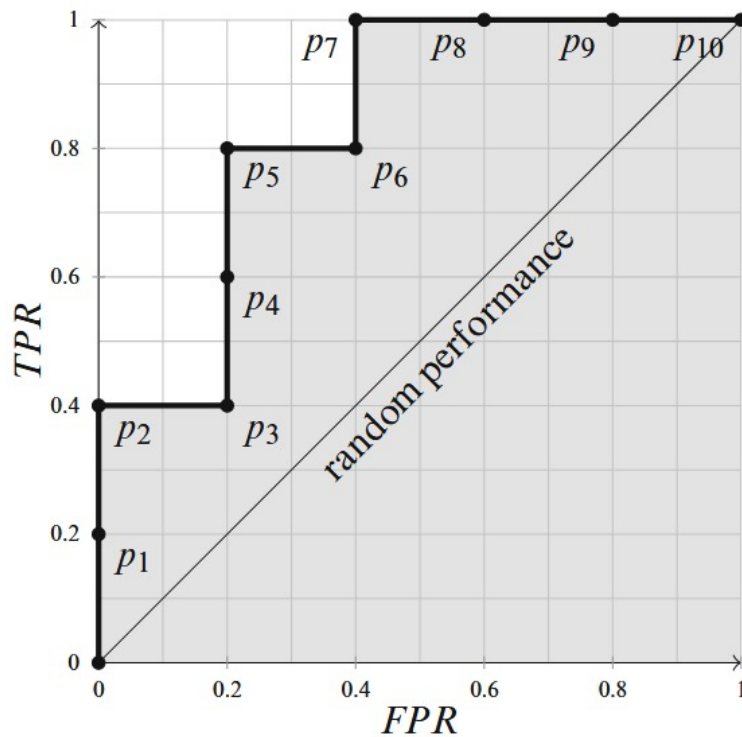


Figura 2.9: Curva de Características do Receptor (Fernandez et al., 2018).

Curva Precision-Recall: Segundo Fernandez et al. (2018), mostra-se mais adequado ao cenário de raridade de amostras do que a AUCROC, por apresentar, de

maneira gráfica, a compensação entre essas duas métricas, tendo o foco apenas na classe positiva. A Figura 2.10 exibe um exemplo de curva *precision-recall*.

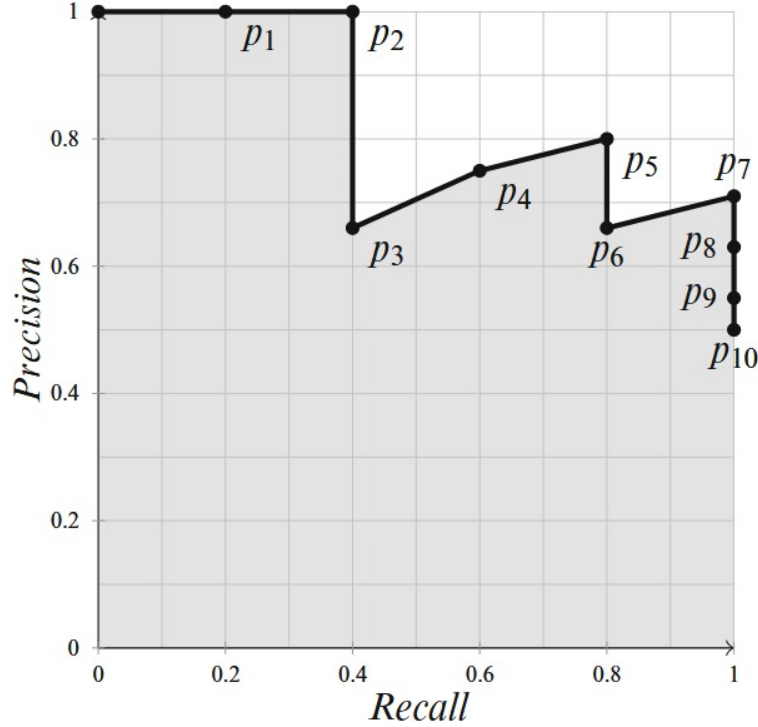


Figura 2.10: Curva Precision-Recall (Fernandez et al., 2018).

Considerando o cenário de acentuado desbalanceamento de algumas classes no conjunto estudado, este trabalho adotou a *F1-score* como principal métrica de avaliação dos modelos. Complementarmente, também foi observada a compensação entre as métricas de *precision* e *recall*. Optou-se por não utilizar a métrica *AUCROC*, devido ao viés que pode ser introduzido em contextos de rara ocorrência de amostras. A acurácia também foi descartada, por não refletir adequadamente o desempenho em cenários desbalanceados.

2.2.8 Few-shot Learning

As técnicas de aprendizado com poucas amostras (FSL) (Fei-Fei et al., 2006) surgiram como estratégias para tratar o cenário de escassez de dados e aliviar o custo operacional de anotação de *datasets* (Wang et al., 2020). A técnica consiste em utilizar o conhecimento prévio obtido com o treinamento em uma grande quantidade de dados rotulados, e em seguida, proceder à realização de diversos pequenos treinamentos, denominados episódios, utilizando uma pequena quantidade de dados rotulados dentro do domínio de interesse. Essa fase é referenciada como abordagem *N-way-K-shot*, em que *N* faz referência à quantidade de classes e *K* refere-se à

quantidade de amostras por classe vistas durante os episódios. O pequeno conjunto de dados rotulados é chamado de conjunto de suporte (*support set*). Na última fase de cada episódio, o modelo aprendido é submetido a um conjunto de consulta (*query set*) para avaliação de seu desempenho (Zhang e Liu, 2022).

Na literatura acadêmica, os métodos mais comuns aplicados em *few-shot learning* são as abordagens baseadas em métrica. Dentre os métodos mais populares, destacam-se as redes siamesas (*siamese networks*) (Koch, 2015), as redes de correspondência (*matching networks*) (Vinyals et al., 2016) e as redes prototípicas (*prototypical networks*) (Snell et al., 2017a), sendo estas últimas as mais populares (Fink, 2004), (Quan et al., 2024), (Parnami e Lee, 2022), (Guan et al., 2020).

O funcionamento das redes siamesas consiste no uso de duas redes convolucionais idênticas com pesos compartilhados, projetadas para comparar pares de imagens de entrada. O treinamento visa ajustar os parâmetros da rede de forma a reduzir a distância entre amostras da mesma classe e aumentá-la entre amostras de classes distintas (Koch, 2015).

Matching networks operam utilizando um mecanismo de atenção para realizar as predições. Para cada episódio, em um primeiro momento, os vetores de características do conjunto de suporte é inteiramente armazenado em memória. A seguir, quando uma amostra do conjunto de consulta é apresentada, uma métrica de similaridade (geralmente a similaridade do cosseno Equação 2.12) é usada para calcular as distâncias entre essa amostra e todos os vetores previamente armazenados. Na última etapa, um mecanismo de atenção é utilizado através da aplicação da função softmax (Equação 2.11) sobre as pontuações de distância obtidas (Vinyals et al., 2016). Na Equação 2.11, z_i são as pontuações de distância calculadas pela rede (*logits*) sobre as N classes; e^{z_i} representa a aplicação da função exponencial sobre o vetor z_i ; $\sum_{j=1}^N e^{z_j}$ é o fator de normalização, aplicado sobre cada elemento z_j contido no vetor z_i , garantindo que a soma de todos os seus elementos seja igual a 1, transformando z_i em uma distribuição de probabilidades.

$$\text{softmax}(z_i) = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (2.11)$$

Nas redes prototípicas o conjunto de suporte é utilizado para a criação dos chamados *protótipos*, que são comparados a novas amostras na tarefa de classificação. Inicialmente, o conjunto de suporte anotado é utilizado durante o treinamento para codificar as amostras em protótipos, calculando a média dos vetores de características extraídos pelas CNNs de *backbone*. Cada classe possui um protótipo, que é armazenado para posterior comparação com as amostras do conjunto de consulta, por meio de uma medida de distância, geralmente a distância euclidiana ou a similaridade do cosseno, cujas fórmulas são apresentadas nas Equações 2.12 e 2.13. Na segunda etapa, um conjunto de consulta contendo novas amostras é submetido ao modelo para validação. Cada amostra recebe a classe do protótipo mais pró-

ximo, de acordo com a menor distância calculada, de maneira análoga ao processo de classificação por vizinhos mais próximos (*K-nearest neighbors* - KNN).

$$\text{cosine_similarity}(A, B) = \frac{A \cdot B}{\|A\| \|B\|} \quad (2.12)$$

$$d(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} \quad (2.13)$$

Segundo Snell et al. (2017a), a técnica consiste no aprendizado de duas funções: uma função codificadora (*embedding function*) f_ϕ , parametrizada por uma rede convolucional com parâmetros ϕ , responsável por transformar as amostras para o espaço de características desejado, e uma função para o cálculo de distâncias (d) entre os vetores de características (*embeddings*) de novas amostras e os protótipos produzidos na primeira etapa.

O treinamento ocorre de forma episódica, com a seleção aleatória de amostras, vindas do conjunto de suporte, seguindo a configuração N -way- K -shot, para a produção dos protótipos, que são testados no conjunto de consulta, em uma quantidade definida de episódios (Quan et al., 2024; Sümer et al., 2023a; Kang Wang e Cheng, 2022; Snell et al., 2017a). Quan et al. (2024) afirmam que abordagens baseadas em métricas são particularmente efetivas no campo da análise de imagens médicas, pois são capazes de capturar diferenças visuais sutis e ainda adaptar-se a cenários de desbalanceamento e escassez de dados. A Figura 2.11 ilustra o funcionamento de uma rede prototípica.

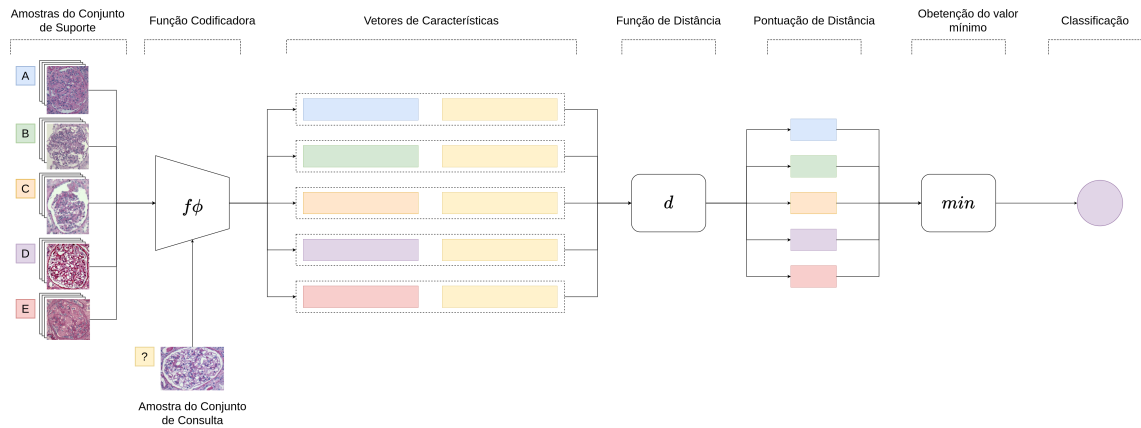


Figura 2.11: Funcionamento de uma Rede Prototípica. As amostras do conjunto de suporte são usadas para a criação dos protótipos com a aplicação da função f_ϕ . Quando uma nova amostra é submetida, esta também é codificada com a função f_ϕ e seu *embedding* é comparado com cada protótipo, gerando uma pontuação de distância com base na função d . O resultado final da classificação é decidido pela cálculo da menor distância ($\min$).

Diferente do que acontece em outras estratégias baseadas em métrica, como as redes siamesas e as redes de correspondência, as redes prototípicas criam uma representação em nível classe utilizando todo o *support set* ao invés de precisar armazenar os vetores de características individuais para comparar com as amostras de consulta. Segundo Snell et al. (2017b), as redes prototípicas, além de serem mais simples e eficientes do ponto de vista computacional - quando comparadas com outras abordagens baseadas em métrica - conseguem alcançar resultados comparáveis às técnicas mais sofisticadas quando se consideram ajustes quanto à configuração dos episódios de treinamento e escolha da função de distância.

2.2.9 Few-shot Learning como estratégia para tratar o Desbalanceamento de Classes

Embora a técnica tenha surgido como uma alternativa para lidar com a escassez de dados, Billion Polak et al. (2024) destacam que as tarefas de FSL representam uma variação desafiadora do problema de classificação em cenários de desbalanceamento. Em ambos os casos, busca-se aprender classes para as quais há poucos dados anotados disponíveis, mas o FSL adiciona a dificuldade adicional da extrema escassez ou raridade de amostras.

Os autores também ressaltam que, apesar da relação entre os dois problemas, uma diferença fundamental está no fato de que os cenários típicos de FSL, geralmente, envolvem uma quantidade extremamente reduzida de amostras, tipicamente entre 1 e 5, tornando inviável o uso dessa abordagem em conjunto com estratégias de balanceamento, como ROS e RUS.

Neste trabalho, optou-se por não empregar técnicas de balanceamento no treinamento das redes prototípicas, pois a introdução de cópias simples das imagens poderia introduzir ruídos na construção dos protótipos, uma vez que estes são calculados a partir da média dos *embeddings* do conjunto de suporte.

A literatura acadêmica recente cita diversos casos de aplicação de técnicas para o tratamento de desbalanceamento usando FSL, seja para o enfrentamento do problema dentro dos treinamentos episódicos ou seja utilizando técnicas de FSL para tratar um desbalanceamento existente (Sümer et al., 2023b; Ochal et al., 2023; Zhu e Yang, 2020).

Ochal et al. (2023) identificaram e formalizaram o conceito de *Class Imbalance Few-Shot Learning* em três níveis de desbalanceamento: (i) *desbalanceamento no nível da tarefa (episódio)*, que pode ocorrer tanto durante o treinamento quanto na avaliação, afetando o conjunto de consulta; (ii) *desbalanceamento no nível do dataset*, que se manifesta quando os episódios são gerados a partir de um conjunto de dados já desbalanceado; e (iii) *desbalanceamento combinado*, que ocorre quando o desbalanceamento está presente tanto no conjunto de dados como em todo o conjunto dentro dos episódios de treinamento. A Figura 2.12 ilustra a diferença entre esses três cenários. É possível ver que, no desbalanceamento no nível do *dataset*, os episódios de

treinamento apresentam uma distribuição balanceada. Nos experimentos realizados neste estudo, envolvendo redes prototípicas, os episódios de treinamento foram configurados de forma a apresentar uma distribuição balanceada dos dados, o processo de validação, entretanto, foi realizado sobre dados com taxa de desbalanceamento equivalente à do conjunto original.

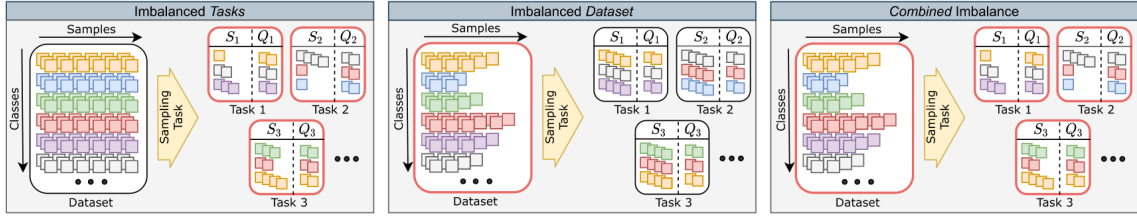


Figura 2.12: Os três níveis de desbalanceamento em CIFSL (Ochal et al., 2023).

2.3 Análise Estatística

A avaliação objetiva e confiável do desempenho de modelos de ML é um elemento crucial para sua validação e eventual aplicação em cenários do mundo real, sendo um tema amplamente debatido na literatura científica (Maier-Hein et al., 2024). Nesse contexto, além da simples comparação entre os resultados obtidos por diferentes modelos, é fundamental verificar se as diferenças observadas apresentam robustez estatística suficiente para sustentar a inferência da superioridade de um modelo sobre outro (Dietterich, 1998). Entre as métricas de significância estatística mais utilizadas para a comparação de modelos de ML, destacam-se (Demšar, 2006):

- **Média sobre diferentes conjuntos de dados:** cálculo da média para uma determinada métrica de desempenho obtida sobre diferentes *datasets*. Segundo Demšar (2006) essa métrica pode ser comprometida se aplicada sobre conjuntos de diferentes domínios, além de ser susceptível à presença de *outliers*;
- **T-Teste pareado (Paired T-Test):** Comumente utilizado para verificar se a diferença entre os desempenhos de dois modelos é consequência de sua aleatoriedade. Consiste basicamente em calcular a média das diferenças observadas em dois modelos distintos, sendo também susceptível a *outliers*, além de assumir uma distribuição normal dos resultados analisados;
- **Teste dos postos com sinais de Wilcoxon (Wilcoxon signed-ranks test):** Considerado uma opção não paramétrica ao t-teste pareado. Considera os valores absolutos das diferenças de desempenho obtidas entre dois modelos. Segundo Demšar (2006), apresenta a vantagem de ser menos influenciado por *outliers*, além de não assumir uma distribuição normal dos resultados.

Neste trabalho optou-se pela utilização da média obtida pelos modelos sobre uma configuração de validação cruzada *K-fold* em conjunto com o *Wilcoxon signed-ranks test*. O teste de Wilcoxon foi escolhido por ser recomendado para a comparação de

dois modelos em contextos de amostras pareadas, especialmente quando se trabalha com um número reduzido de resultados, como é o caso dos cinco *folds* utilizados neste estudo, além de ser robusto à presença de *outliers* e não exigir a suposição de normalidade na distribuição dos dados (Demšar, 2006).

Neste capítulo foram apresentados os principais fundamentos deste trabalho, introduzindo tópicos da Patologia e Nefropatologia até a Computação. Foram explicados os conceitos de Histologia e o processo de aquisição de imagens para o treinamento de modelos de ML, incluindo o papel das redes convolucionais nas tarefas de classificação dessas imagens. Por fim, foram apresentados os conceitos de desbalanceamento de classes, as principais abordagens para o seu enfrentamento e o uso de técnicas como FSL em tais cenários. O próximo capítulo apresenta os principais trabalhos relacionados com esta pesquisa, descrevendo seus contextos, estratégias e resultados alcançados.

Capítulo 3

Trabalhos Relacionados

Este capítulo apresenta um conjunto de trabalhos relacionados com o tema desta pesquisa., visando dar um panorama sobre técnicas recentemente utilizadas para solucionar problemas similares aos aqui enfrentados. Também intenciona estabelecer um contexto em relação ao desempenho dos métodos utilizados, que será útil na análise dos resultados obtidos neste trabalho.

3.1 Trabalhos que utilizaram Técnicas Comuns para tratar o Desbalanceamento de Classes

As técnicas mais comuns para o tratamento de cenários de desbalanceamento podem ser aplicadas de forma isolada ou combinada. Nesta seção, são apresentados estudos que empregam essas abordagens amplamente difundidas, destacando sua popularidade e consolidação na literatura científica. Embora os trabalhos selecionados nem sempre apresentem propostas inovadoras, eles reforçam a relevância e a recorrência do uso dessas estratégias em diferentes contextos. Além disso, foram priorizados estudos voltados à patologia computacional, especialmente no que se refere aos resultados obtidos, os quais serviram como referência para o estabelecimento de uma meta de desempenho neste trabalho.

Mahbub et al. (2024) empregaram um método algorítmico e propuseram uma nova função de custo, *Perda de Afinidade com Foco Central* (*Center-Focused Affinity Loss* - CFAL), que altera o comportamento da função de ativação da última camada da rede, introduzindo sensibilidade ao custo por meio de uma função de perda afinidade, mitigando os efeitos da variação intra-classe e aumentando a distância inter-classe. Os experimentos foram realizados em *datasets* histológicos e alcançaram até 83% para F1-score em um conjunto com IR de $\approx 1 : 3$.

Walsh e Tardy (2023) realizaram um comparativo entre técnicas comuns para desbalanceamento e o uso de *redes adversariais generativas* (*Generative Adversarial*

Networks - GAN) para a classificação de imagens de mamografias usando três *datasets* distintos. Notadamente, foram empregados no estudo os métodos RUS, ROS e CSL. Para evitar o potencial *overfitting* introduzido pela simples geração de cópias das amostras, foram utilizadas GANs, seguindo o modelo proposto por Tremblay et al. (2018). Os autores denominaram o método proposto como *Artifacting* e para o conjunto com a maior taxa de desbalanceamento, na ordem de 1:19, contendo 20.000 imagens, os autores alcançaram uma taxa de 76,8% para AUCROC.

Raj et al. (2022) propuseram uma nova técnica de *data augmentation*, denominada *Técnica baseada em Cruzamento (Crossover-based Technique)*, que combina duas imagens existentes para gerar novas amostras, utilizando o operador de *crossover* dos algoritmos genéticos. Nessa abordagem, a seleção das amostras a serem combinadas ocorre a partir de um *mating pool*, composto pelos indivíduos selecionados com base em critérios de aptidão (*fitness*), em um processo análogo à *seleção natural*. Os autores realizaram o treinamento com os dados aumentados utilizando CNNs, com três *datasets* médicos, chegando a alcançar um Macro F1-Score de 98% em um conjunto multiclasse para a detecção de tumores cerebrais. A distribuição do conjunto é dada na Figura 3.1.

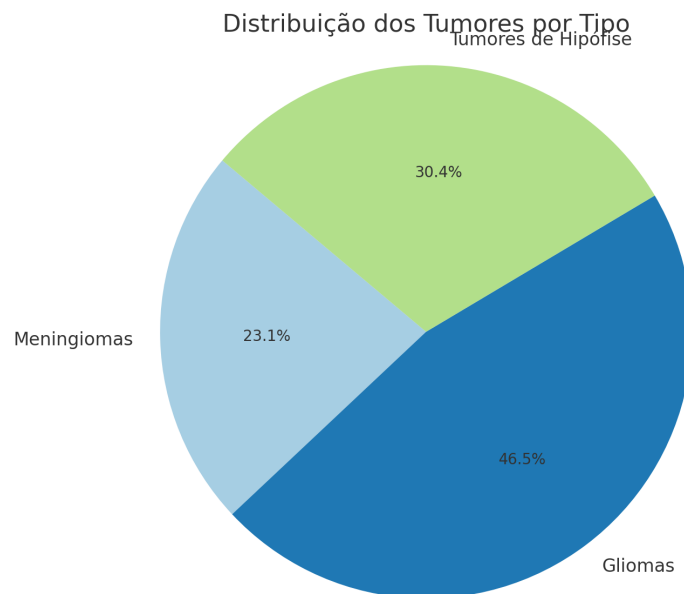


Figura 3.1: Distribuição do conjunto utilizado por Raj et al. (2022). O *dataset* é composto por 3.064 imagens.

Song et al. (2021) usaram CNNs em conjunto com métodos baseados em dados e CSL para a classificação automática de câncer oral, considerando a classe normal e três classes representando níveis de malignidade. A distribuição dos dados é dada na Figura 3.2. Os autores alcançaram um *Macro F1* de 81%.

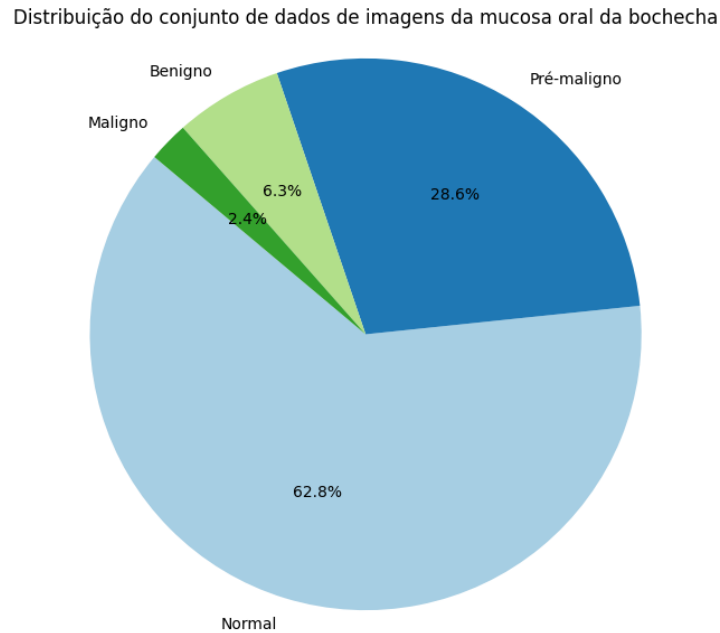


Figura 3.2: Distribuição do conjunto utilizado por Song et al. (2021). O *dataset* é composto por 3.851 amostras. Fonte: o próprio autor.

Chen et al. (2024b) empregaram técnicas de *data augmentation* e reamostragem em uma tarefa de classificação multiclasse, combinando *datasets* para reconhecimento de expressões faciais (*Facial Expression Recognition* - FER), cuja distribuição é apresentada na Figura 3.3. Os objetivos do trabalho foram: (i) abordar a inconsistência nas anotações resultante da combinação de diferentes conjuntos de dados e (ii) demonstrar a influência do desbalanceamento de classes nesse processo. Para isso, os autores propuseram um novo método de treinamento com junção de *datasets*, denominado *Seleção de Amostras e Treinamento com Aumento Pareado* (*Sample Selection and Paired Augmentation Joint Training* - SSPA-JT). Nesse método, apenas as amostras pertencentes às classes minoritárias são, de forma algorítmica, selecionadas do conjunto auxiliar e incorporadas ao conjunto-alvo, com o objetivo de ampliar sua representatividade. O algoritmo também trata da busca por imagens com anotação considerada consistente, tomando como referência amostras previamente selecionadas e marcadas como *seguras*. Os experimentos mostraram que a simples fusão completa dos conjuntos impactou negativamente o desempenho dos modelos, ao passo que a combinação seletiva de amostras minoritárias resultou em um ganho de desempenho de até 1,15%. O estudo alcançou uma taxa de *Macro F1* de até 87,66%.

Ilmawati et al. (2024) Empregaram técnicas de DA em conjunto métodos híbridos de reamostragem, empregando SMOTE e sub-amostragem, utilizando a CNN VGG-19 para a classificação de expressões faciais, em um dos conjuntos FERF (Dumitru, 2013), cuja distribuição é dada na Figura 3.4. Os autores alcançaram um *Macro F1* de 66,58%.

Nourivandi et al. (2024) empregaram CNNs e SMOTE para o reconhecimento de expressões faciais. O modelo produzido foi avaliado com o *dataset* AffectNet (Mollahosseini et al., 2019), cuja distribuição é dada na Figura 3.3(c), alcançando um *Macro F1* de 62%.

Dai et al. (2022) fizeram uso de *transfer learning* e técnicas de DA por meio de redes adversariais generativas para a classificação de pontos de solda defeituosos. A distribuição do conjunto, composto por 9.214 imagens, é dada na Figura 3.5. O trabalho alcançou um desempenho de 98,1% para *Macro F1*.

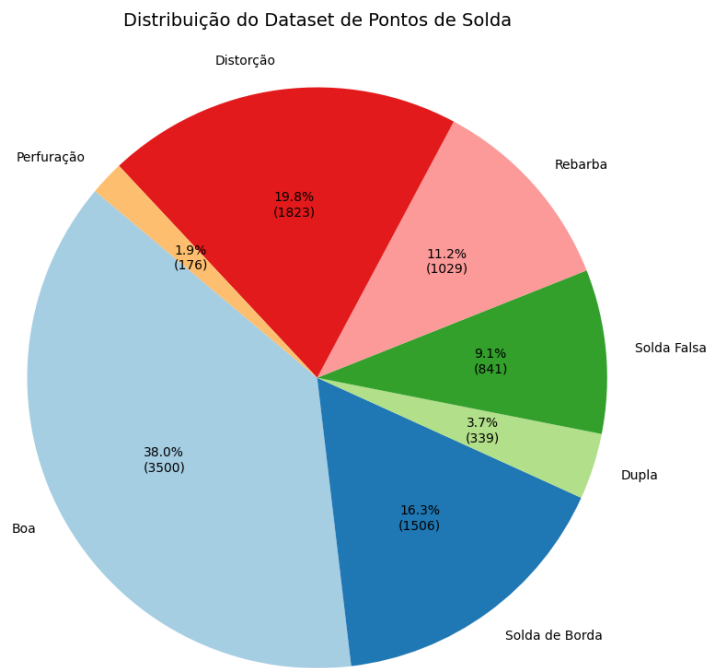


Figura 3.5: Distribuição do conjunto utilizado por Dai et al. (2022). Fonte: o próprio autor.

Haque et al. (2022) usaram *data augmentation* e CNNs para uma tarefa de classificação de doenças em plantações de milho. A distribuição do conjunto, composto de 5.939 imagens, é dada na Figura 3.6. Os autores obtiveram um *Macro F1* de 95,66%.

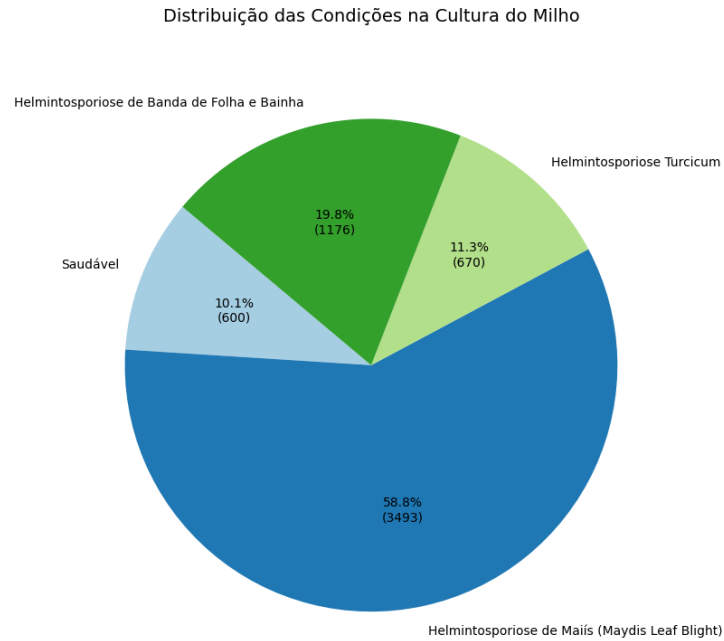


Figura 3.6: Distribuição do conjunto utilizado por Haque et al. (2022).

Duarte et al. (2020) fizeram uso de ROS, CSL e empregaram um *limiar de classificação*¹ para tratar um desbalanceamento de 1:100 em um conjunto para classificação de imagens contendo lixo marinho, composto por 54.022 amostras. A técnica ROS alcançou o maior F1-score, chegando a 68%. A Tabela 3.1 exibe os resultados alcançados no campo da patologia computacional, enquanto a Tabela 3.2 exibe os resultados alcançados em outros domínios.

Tabela 3.1: Sumário das técnicas e resultados para os trabalhos realizados na patologia computacional

Estudo	Técnica(s)	IR	Métrica	Resultado
Mahbub et al. (2024)	CFAL	$\approx 1:3$	F1-Score	83,00%
Walsh e Tardy (2023)	<i>Artifacting</i>	$\approx 1:19$	AUCROC	78,60%
Raj et al. (2022)	Crossover DA	—	M. F1	98,00%
Song et al. (2021)	ROS+CSL-CNN	—	M. F1	81,00%

¹Esta técnica é implementada pela escolha de um limite a partir do qual um modelo considera uma amostra como pertencente a uma determinada classe.

Tabela 3.2: Sumário das técnicas e resultados para os trabalhos relacionados a terem empregado métodos comuns fora da patologia computacional

Estudo	Técnica(s)	IR	Métrica	Resultado
Chen et al. (2024b)	SSPA-JT	—	M. F1	87,66%
Ilmawati et al. (2024)	DA+SMOTE+US	—	M. F1	66,58%
Nourivandi et al. (2024)	CNN+SMOTE	—	M. F1	62,00%
Dai et al. (2022)	TL+GAN	—	M. F1	98,10%
Haque et al. (2022)	DA+CNN	—	M. F1	95,66%
Duarte et al. (2020)	ROS	1:100	F1-Score	68,00%

3.2 Trabalhos que utilizaram Few-shot Learning e CIFSL

O uso mais convencional de FSL é para o regime de escassez de dados, contudo são encontradas também referências de sua utilização em cenários de desbalanceamento. Nesse contexto, a técnica é também citada como *Class Imbalance Few-shot Learning* (Ochal et al., 2023). Para esta seção foram selecionados trabalhos em que a técnica foi notadamente utilizada para enfrentar o desbalanceamento de classes, isoladamente ou em conjunto com métodos comuns, demonstrando a sua utilização em domínios diversos, com ênfase na sua utilização na patologia computacional.

Deng e Li (2023) utilizaram TL, FSL, técnicas de reamostragem e métodos de mascaramento de imagens (*image masking*) para a classificação e contagem de glóbulos brancos em amostras sanguíneas com desbalanceamento da ordem de 1:6. *Image masking* é uma técnica utilizada para omitir *pixels* de uma imagem que encontram-se fora da área de interesse para uma determinada tarefa (Xie et al., 2022). Os autores obtiveram resultados de até 88% para AUCROC, utilizando apenas o mascaramento de imagens na etapa de pré-processamento.

Abbas (2023) compararam o desempenho de três abordagens de FSL com o de uma *deep CNN* tradicional em um sistema de recuperação de imagens. As técnicas de *few-shot learning* avaliadas incluíram redes prototípicas, redes siamesas e o uso de uma *deep CNN* pré-treinada como rede *backbone*, empregada para aprender representações a partir de dados aumentados por meio de DA e posteriormente utilizadas na construção de protótipos. Os experimentos foram conduzidos sobre os conjuntos TCIA (Clark et al., 2013) e KVASIR (Pogorelov et al., 2017), alcançando *sensibilidade* de 95%, *especificidade* de 96,5%, AUC de 0,96 e *acurácia* de 97,5%.

Tituriya e Singh (2022) usaram redes prototípicas e meta aprendizado agnóstico de modelo (*Model Agnostic Meta Learning* - MAML) em quatro diferentes *datasets* para o diagnóstico de cancer. O trabalho apresenta resultados de até 84,56% de acurácia para uma configuração 2-way-2-shot no conjunto de consulta. A distribuição do conjunto é dada na Figura 3.7.

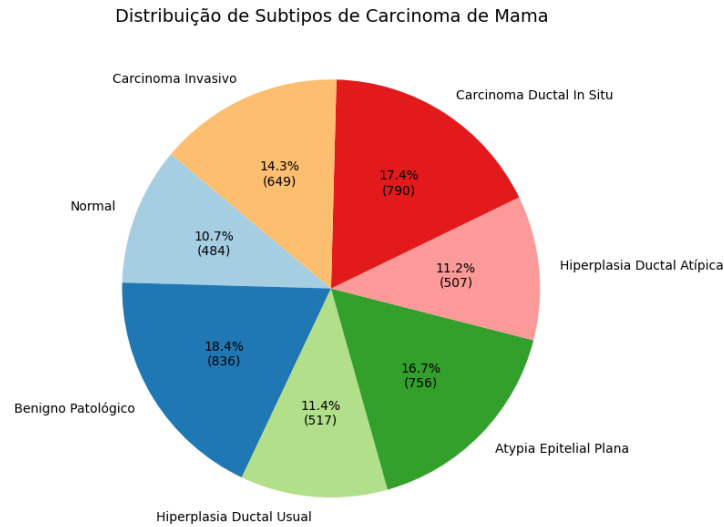


Figura 3.7: Distribuição do conjunto utilizado por Titoriya e Singh (2022).

Ochal et al. (2023) realizaram um extenso estudo, utilizando FSL em conjunto com técnicas de balanceamento e CSL. Os autores focaram prioritariamente no tratamento do desbalanceamento ocorrido dentro dos episódios de treinamento, realizando tanto episódios N-way-K-shot típicos, quanto aplicando *random-shot meta-training*, em que a quantidade de amostras selecionadas do *support-set* para cada classe, é definida de forma aleatória, introduzindo um desbalanceamento no nível de tarefa, utilizando as configurações N-way-(4-6)-shot e N-way-(1-9)-shot. Os experimentos foram conduzidos sobre o conjunto MiniImageNet (Vinyals et al., 2017) e foi observado que o desempenho dos modelos foi negativamente impactado quando estes foram avaliados em conjuntos de consulta desbalanceados. Para a configuração N-way-(1-9)-shot, foi alcançada uma taxa de 62,09% nos testes com dados balanceados e 61,44% com o conjunto de consulta desbalanceado, com precisão de 42% e *recall* de 9% para a classe minoritária. No presente estudo, todas as avaliações dos modelos construídos com FSL foram realizadas sobre dados desbalanceados.

Guan et al. (2020) propuseram um novo framework para o treinamento Episódico em FSL, composto por treinamento randômico aleatório para as classes majoritárias e uma etapa adicional de ajuste fino (*fine-tuning*) com todas as classes; o método foi testado com sub-conjuntos derivados dos *datasets* NPWU-RESISC45 e AID, com desbalanceamentos nas proporções 1:10, 1:14, 1:20 e 1:28. Os autores alcançaram resultados de até 75,51% para acurácia balanceada, na maior taxa de desbalanceamento.

A Tabela 3.3 apresenta um sumário, contendo as técnicas utilizadas, a taxa de desbalanceamento (IR) (quando aplicável) do *dataset*, as métricas de desempenho e o resultado obtido para os estudos que utilizaram FSL no campo da patologia computacional, enquanto a Tabela 3.4 lista os resultados obtidos em outras áreas.

Tabela 3.3: Sumário dos resultados para FSL na patologia computacional.

Estudo	Técnica(s)	IR	Métrica	Resultado
Deng e Li (2023)	TL + Image Mask	1 $\approx$:6	AUCROC	88,00%
Abbas (2023)	FSL + DA	-	Acurácia	97,50%
Titoriya e Singh (2022)	PN+MAML	-	Acurácia	84,56%

Tabela 3.4: Sumário dos resultados para FSL fora da patologia computacional.

Estudo	Técnica(s)	IR	Métrica	Resultado
Ochal et al. (2023)	FSL+ROS+CSL	1:9	BAC	62,09%
Guan et al. (2020)	RF-MML-SVM	1:28	BAC	75,51%

3.3 Análise dos Trabalhos Relacionados

Esta seção apresenta as principais observações feitas com respeito aos trabalhos relacionados a este estudo, concernente às técnicas empregadas, dados utilizados, taxa de desbalanceamento e resultados alcançados. Vale registrar que após uma extensa revisão da literatura sobre classificação na Patologia Computacional, não foi identificado qualquer estudo que integre um processo de curadoria do conjunto de dados e o treinamento de modelos de ML com amostras restritas a corantes específicos. Também não foram encontradas referências de trabalhos com a profundidade aqui apresentada no tocante à classificação de lesões glomerulares. Tais estudos serviriam como elementos de comparação para os resultados obtidos neste trabalho.

3.3.1 Análise dos trabalhos com Técnicas Comuns

Considerando os estudos citados neste trabalho, dentre aqueles que abordaram tarefas de classificação binária, dois adotaram F1-score como métrica de referência enquanto um outro trabalho adotou a AUCROC como métrica principal. Mahbub et al. (2024) e Duarte et al. (2020) utilizaram métodos diferentes para enfrentar o desbalanceamento. Enquanto Mahbub et al. (2024) focaram em uma abordagem totalmente algorítmica, propondo uma nova função de custo para alavancar o desempenho dos modelos e aplicaram CSL, Duarte et al. (2020) optaram por estratégias baseadas em dados, combinando-as em alguns experimentos com a técnica de *classifier threshold*.

Walsh e Tardy (2023) adotaram a AUCROC como métrica em um cenário com IR de 1:19, alcançando um desempenho de 78,6% para essa métrica. Porém, os resultados obtidos pelos autores em outras métricas, especialmente para uma métrica mais robusta para desbalanceamento, o MCC, denotam a fragilidade da escolha de AUCROC em casos de desbalanceamento mais acentuado (Johnson e Khoshgoftaar, 2019). O mesmo modelo obteve os valores 49,9% para sensibilidade, 87,3% para

especificidade e apenas 25,3% para MCC. O valor relativamente alto para *especificidade* realça o viés do modelo para os verdadeiros negativos, indicando um melhor desempenho para a classe majoritária.

Para os trabalhos voltados para tarefas multiclasse, três estudos alcançaram um *Macro F1* acima de 95%. Nos três trabalhos, foram utilizadas estratégias baseadas em dados, tendo Dai et al. (2022) e Raj et al. (2022) utilizado métodos baseados em GAN, enquanto Haque et al. (2022) empregou DA. Dai et al. (2022) obteve o melhor resultado em um conjunto com sete classes, em contraste com os outros dois trabalhos que continham apenas quatro classes. Para cada conjunto utilizado, considerando-se somente as classes minoritárias e majoritárias, encontramos uma relação para Dai et al. (2022) de $\approx 1:20$, para Raj et al. (2022) de $\approx 1:4$, e para Haque et al. (2022) de $\approx 1:6$. É possível que a complexidade dos dados tenha contribuído para o desempenho robusto desses classificadores.

Nas demais tarefas multiclasse, Chen et al. (2024b), Nourivandi et al. (2024) e Ilmawati et al. (2024) exploraram o mesmo problema, o reconhecimento de expressões faciais. Todos os autores abordaram o problema com estratégias baseadas em dados, notadamente SMOTE e DA. No trabalho realizado por Nourivandi et al. (2024), SMOTE foi aplicado aos *vetores de características* (*embeddings*) extraídos por uma CNN, enquanto Ilmawati et al. (2024) aplicou a técnica em conjunto com DA e adicionou uma etapa de sub-amostragem para eliminação de ruídos. Chen et al. (2024b), empregaram um treinamento com a junção de três conjuntos de dados, em que para cada conjunto utilizado para treinamento, cópias para as classes minoritárias eram trazidas dos outros conjuntos. O aumento no desempenho obtido nesse trabalho, pode ser justificado pelo fato de o aumento de dados ser feito com o auxílio de amostras orgânicas de conjuntos auxiliares, ao invés da replicação sintética de imagens já contidas em um conjunto, o que pode ter introduzido uma maior variedade de dados, contribuindo para uma melhor generalização do modelo construído.

Song et al. (2021) aplicaram ROS e CSL para a classificação de câncer bucal, combinando métodos baseados em dados e técnicas algorítmicas. O estudo foi realizado com 3.851 amostras, das quais, apenas 91 pertenciam à classe minoritária, com 2.417 imagens para a classe majoritária. Os autores alcançaram um Macro F1-score de 81%, com um F1 para a classe minoritária (maligno) de 84% e 95% para a classe majoritária (normal). O F1 mais baixo foi registrado para tumores benignos, com 68%. Observando ainda os resultados para a classe minoritária, foram alcançados valores para *revocação* (*recall*) 75% e precisão 95%, indicando que o modelo produzido teve poucos falsos positivos, mas apresentou um número relativamente alto de falsos negativos.

3.3.2 Análise dos trabalhos que fizeram uso de FSL e CIFSL

Dentre os estudos que empregaram FSL, Medela et al. (2019) e Abbas (2023) realizaram trabalhos fora do contexto de desbalanceamento, porém, foram incluídos neste estudo para apresentação da técnica FSL no cenário de escassez de dados,

no qual a técnica foi originalmente concebida. Embora os conjuntos utilizados nos dois estudos não sejam necessariamente escassos, o treinamento episódico realizado por esses trabalhos, utilizou uma quantidade limitada de amostras no conjunto de suporte, o que ilustra a principal motivação para uso da técnica.

Ochal et al. (2023) não realizaram um estudo com objetivo de alcançar ou superar alguma referência de desempenho. Antes, os autores se propuseram a demonstrar os desafios encontrados em diferentes cenários ou níveis de desbalanceamento em cenários tratados com CIFSL. Em suas observações, fica evidenciado que o desbalanceamento em nível de tarefa/episódio apresenta maiores entraves ao desempenho dos classificadores, do que o desbalanceamento em nível de *dataset*.

Medela et al. (2019) realizaram treinamentos episódicos com um número de amostras que variou de 1 a 100, alcançando 90% de acurácia, o que em um contexto sem desbalanceamento, mostra-se válido. Abbas (2023) realizaram treinamentos com 5-way e 15-way com uma quantidade amostras (*shots*) variando de 1 a 15. Ambos os trabalhos utilizaram abordagens baseadas em métricas. Em nenhum dos dois estudos se observou alguma validação feita com dados externos aos conjuntos de consulta (*query sets*). Titoriya e Singh (2022) também realizaram treinamentos episódicos com amostras limitadas, porém, o estudo utilizou quatro diferentes *datasets*, dois dos quais, apresentavam desbalanceamento. Os autores também utilizaram métodos baseados em métrica, e não validaram o modelo produzido sob condições de desbalanceamento, atendo-se somente à validação realizada nos *query sets*, dentro dos treinamentos episódicos balanceados.

Deng e Li (2023) e Guan et al. (2020) abordaram tarefas com CIFSL, empregando abordagens baseadas em métricas para cenários de desbalanceamento. Embora a IR presente no trabalho realizado por Deng e Li (2023) possa ser considerada moderada (1:6), o uso da métrica AUCROC não oferece um aferidor confiável, por conta de seu comportamento tornar-se menos sensível aos falsos positivos à medida que quantidade de amostras negativas cresce, apresentando resultados muito otimistas (Johnson e Khoshgoftaar, 2019). Os autores combinaram o mascaramento de imagens com técnicas de reamostragem, alcançando uma taxa de 88%. Guan et al. (2020) realizaram o treinamento episódico com dados desbalanceados, configurando os episódios de forma a selecionar um número limitado de amostras das classes minoritárias, enquanto preservando quantidade de amostras majoritárias. Apesar do estudo ter alcançado uma BAC de 75,51%, o desempenho para a classe minoritária foi muito inferior ao alcançado para a majoritária, 50,99% contra 87,87%.

Este capítulo apresentou os trabalhos considerados correlatos ao estudo aqui desenvolvido. A partir da análise dessas pesquisas, observa-se a eficácia de abordagens algorítmicas, de métodos baseados em dados, e também da combinação entre ambas na construção de modelos em cenários de desbalanceamento. O uso de *few-shot learning* nesses contextos é também evidenciado pelos autores citados, o que reforça a pertinência dessa abordagem, especialmente em situações de escassez de amostras. Com base nessas evidências, este trabalho adotou uma estratégia híbrida,

combinando métodos tradicionais e FSL com o aprendizado sensível ao custo, com o objetivo de aprimorar o desempenho dos classificadores por meio da sinergia entre técnicas algorítmicas e orientadas por dados.

O próximo capítulo apresenta a metodologia aplicada neste estudo, detalhando os procedimentos e estratégias empregados na construção dos modelos e os critérios e métricas de avaliação.

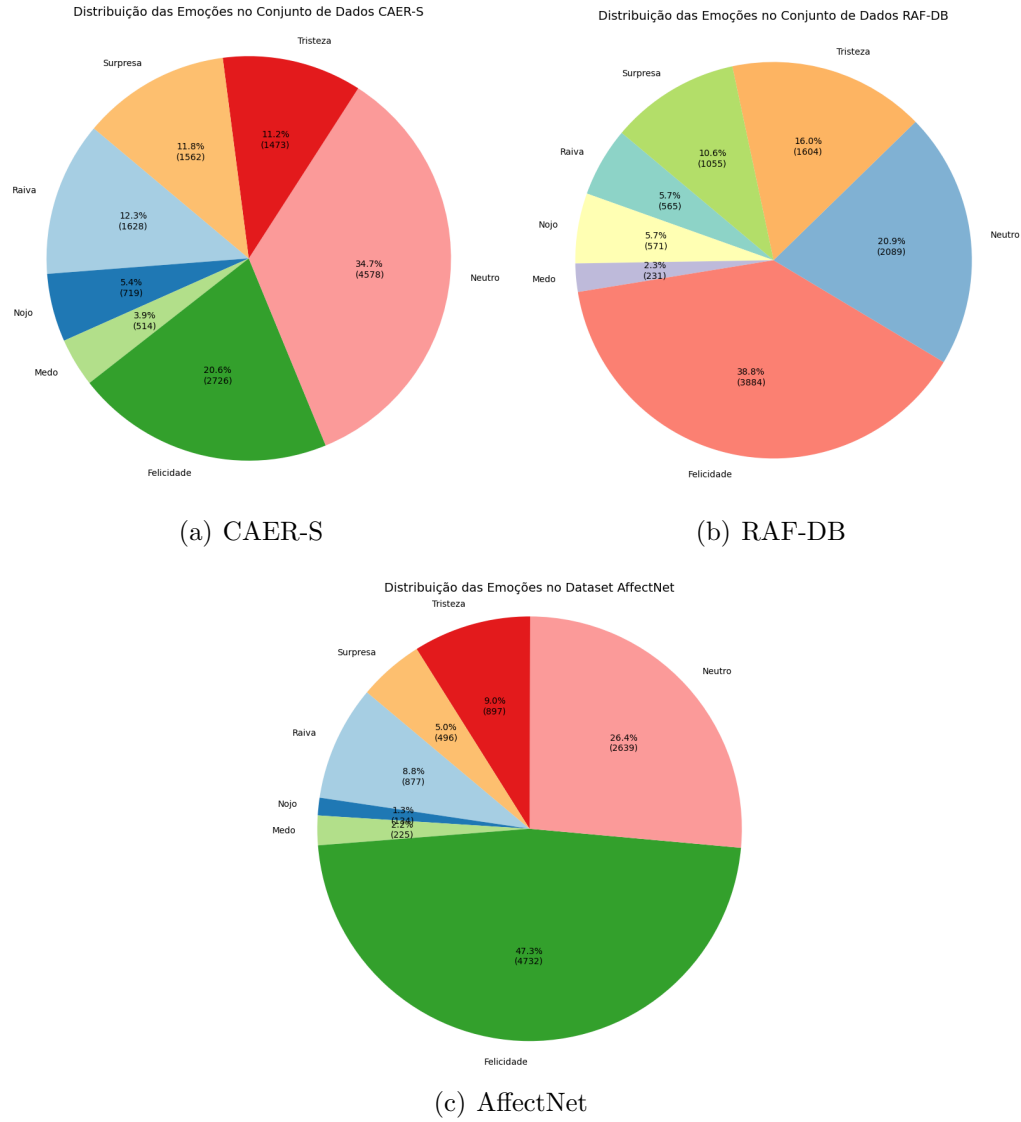


Figura 3.3: Distribuição dos *datasets* empregados no estudo realizado por Chen et al. (2024b). Os conjuntos CAER-S, RAF-DB e AffectNet utilizados no estudo possuem 65.983, 15.339 e 283.721 amostras respectivamente. Fonte: o próprio autor.

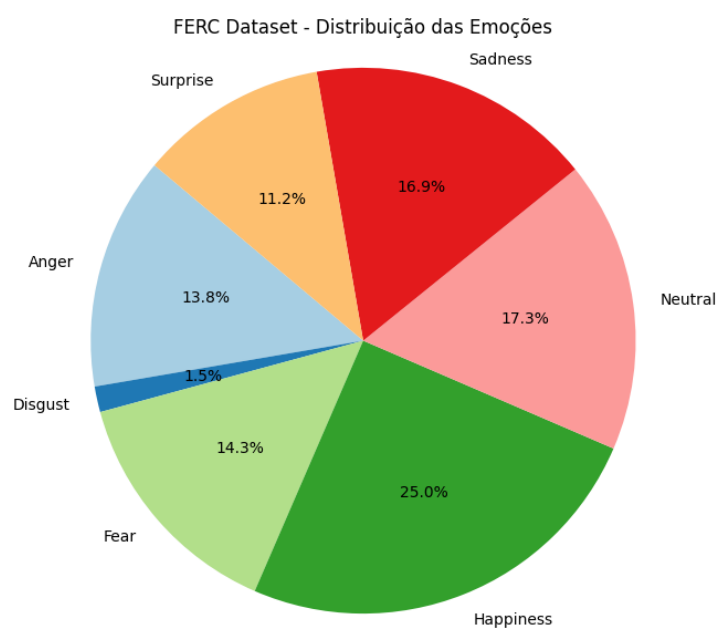


Figura 3.4: Distribuição do conjunto utilizado por Ilmawati et al. (2024). O conjunto possui um total de 35.887 imagens. Fonte: o próprio autor.

Capítulo 4

Metodologia

Este capítulo apresenta o conjunto de dados, as técnicas e estratégias utilizados neste estudo, com uma descrição detalhada do processo de construção dos modelos para classificação binária. Primeiro é apresentado o conjunto de dados original e as escolhas feitas para adequá-lo ao treinamento dos classificadores, que incluíram a criação de conjuntos segmentados por corante. Para avaliar a influência da qualidade das amostras nos resultados dos experimentos, o *dataset* foi submetido a um processo de curadoria, no qual se procurou mitigar potenciais ruídos, introduzidos por amostras consideradas fora do padrão para os experimentos. Detalhes sobre os critérios utilizados no processo de curadoria são descritos nas seções a seguir.

Os experimentos aconteceram com a criação de diversos modelos, construídos com as abordagens mais comuns para o desbalanceamento de classes encontradas na literatura como reamostragem e aprendizado sensível ao custo. Foram construídos classificadores binários para todas as classes, utilizando-se a técnica de decomposição *one versus all* descrita no Capítulo 2. Esses modelos foram construídos para que se escolhesse um *baseline* para fins comparativos. As tarefas realizadas utilizaram conjuntos com amostras em todos os corantes e classes, bem como conjuntos com restrições quanto a corante ou classe. Especificamente, foram selecionadas amostras tingidas com HE e PAS, por serem estes os mais representados no conjunto utilizado neste trabalho.

Após a construção dos modelos *baseline*, procedeu-se ao desenvolvimento de classificadores com *Few-shot Learning* através de tarefas de classificação binária, também com dados contendo todas as classes e corantes e com conjuntos limitados à presença dos corantes HE e PAS. O corante HE é citado na literatura como o mais largamente utilizado na coloração histológica (Montanari, 2016), sendo também o mais presente no *dataset* aqui utilizado. Já o corante PAS foi escolhido por ser o único a apresentar amostras de todas as classes contidas no conjunto original. Por fim, foi adotada a combinação de FSL com a técnica de aprendizado sensível ao custo.

4.1 O Conjunto de Dados Original

Este trabalho utilizou um dataset fornecido pelo Dr. Washington Luis Conrado dos Santos, do Instituto Gonçalo Moniz da Fundação Oswaldo Cruz - Bahia, contendo 11.674 amostras de glomérulos extraídas de imagens de lâmina completa, no formato JPG, distribuídas entre amostras sem lesão, amostras com a lesão amiloidose, amostras com esclerose pura sem crescente, amostras com lesão do tipo hiper celularidade, amostras com lesão do tipo hiper celularidade pura sem crescente, amostras com glomerulonefrite crescente, amostras com nefropatia membranosa, amostras com esclerose e amostras com podocitopatia, obtidas nos corantes Tricromo de AZAN (AZAN), Hematoxilina e Eosina (HE), Ácido Periódico-prata Metenamina (*Periodic Acid-Methenamine Silver*) - PAMS) e Ácido Periódico de Schiff (*Periodic Acid-Schiff* - PAS). Quanto às dimensões das imagens utilizadas, existe certa diversidade, com tamanhos variando de 607px $\times$ 751px a 1024px $\times$ 768px, todas com resolução de 300 $\times$ 300 ppp. A distribuição das amostras é descrita nas Tabelas 4.1 e 4.2. As Figuras 4.1 a 4.3 exibem imagens de glomérulos com as diferentes lesões e corantes utilizados neste estudo.

Tabela 4.1: Distribuição do conjunto de dados original por lesão e corante.

Classe	AZAN	HE	PAMS	PAS	Total
Amiloidose	31	145	96	102	374
Normal	223	1.585	345	542	2.695
Esclerose Pura	234	672	104	472	1.482
Hiper celularidade	257	1.890	0	987	3.134
Hiper celularidade Pura	60	0	0	164	224
Glomerulonefrite Crescente	121	467	157	359	1.104
Nefropatia Membranosa	136	712	324	367	1.539
Esclerose	0	276	122	219	617
Podocitopatia	90	65	106	244	505
Total	1.152	5.812	1.254	3.456	11.674

Tabela 4.2: Distribuição do conjunto de dados original por lesão e corante em valores percentuais.

Classe	AZAN	HE	PAMS	PAS	Total
Amiloidose	0,2%	1,2%	0,8%	0,8%	3%
Normal	1,9%	13,58%	2,9%	4,6%	23%
Esclerose Pura	2%	5,7%	0,9%	4%	12,6%
Hiper celularidade	2,2%	16,2%	-	8,4%	26,8%
Hiper celularidade Pura	0,5%	-	-	1,4%	1,9%
Glom. Crescente	1%	4%	1,3%	3,1%	9,4%
Nefropatia Membranosa	1,2%	6%	2,8%	3,1%	13,1%
Esclerose	-	2,4%	1%	1,9%	5,3%
Podocitopatia	0,8%	0,5%	0,9%	2%	4,2%
Total	9,8%	49,58%	10,6%	29,3%	100%

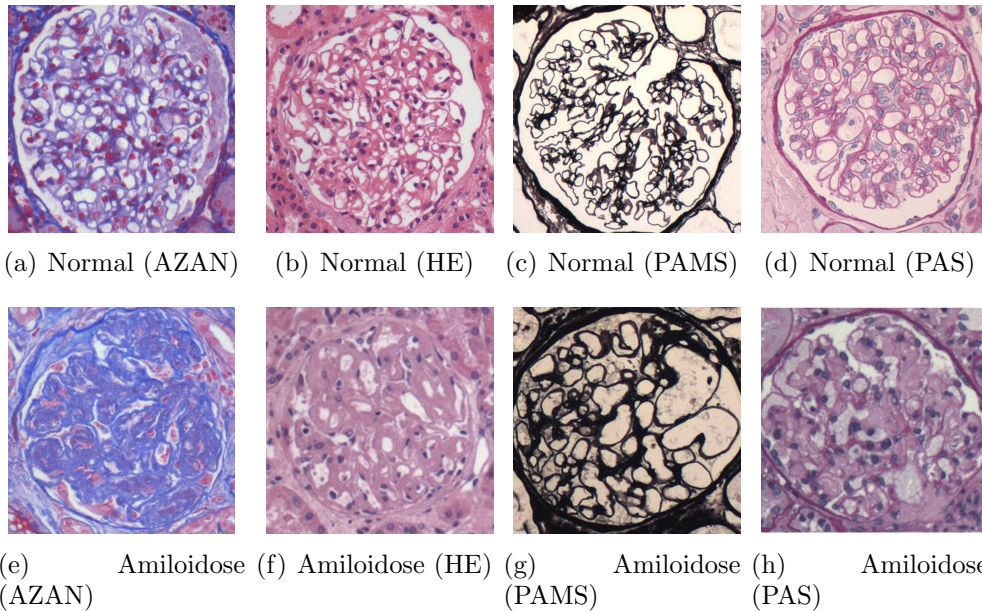


Figura 4.1: Imagens de glomérulos presentes no conjunto de dados. Da esquerda para a direita, são apresentadas amostras tingidas respectivamente com os corantes AZAN, HE, PAMS e PAS. As Figuras 4.1(a) a 4.1(d) apresentam glomérulos sem lesão; 4.1(e) a 4.1(h), glomérulos com *amiloidose*.

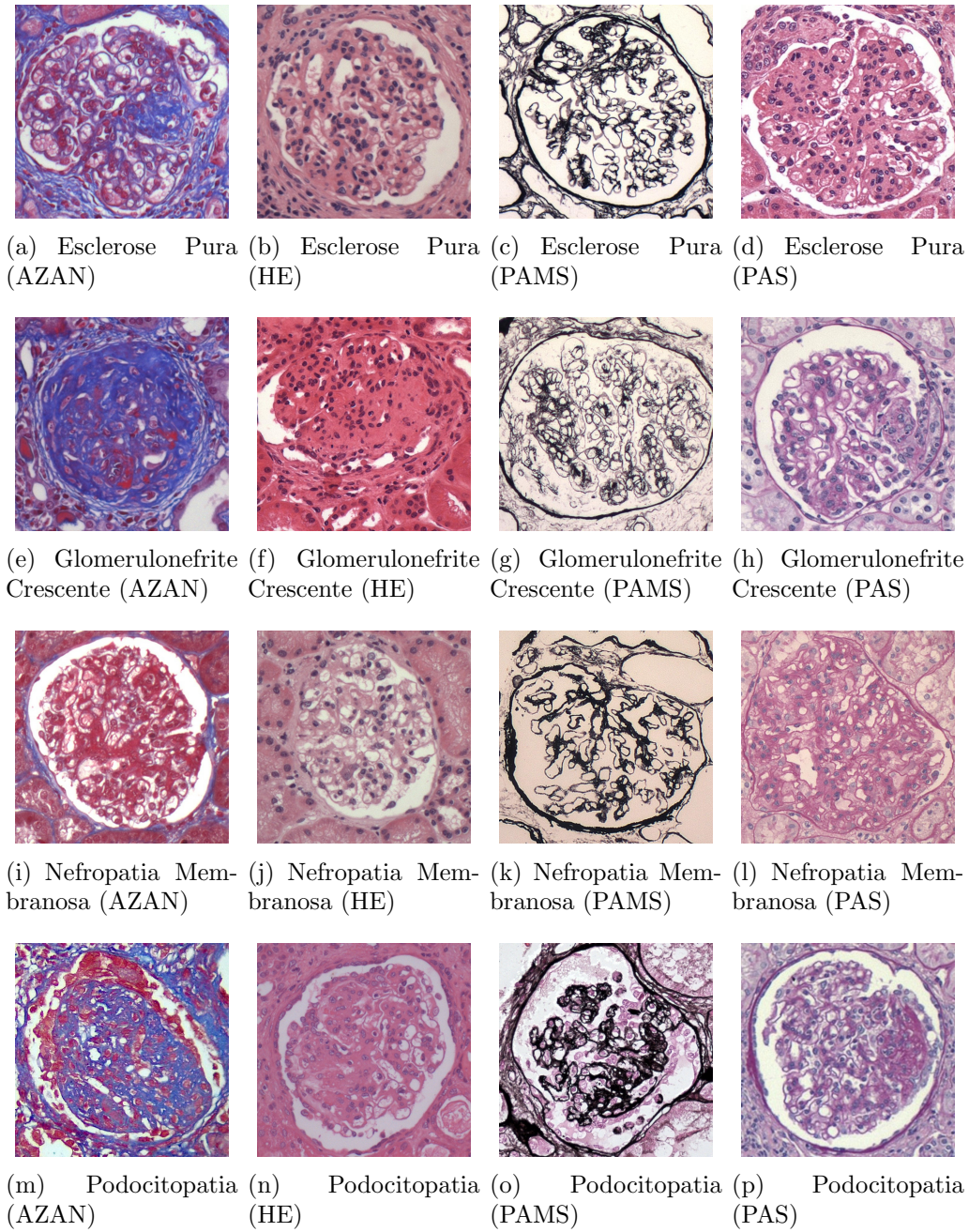


Figura 4.2: Imagens de glomérulos presentes no conjunto de dados. Da esquerda para a direita, são apresentadas amostras tingidas respectivamente com os corantes AZAN, HE, PAMS e PAS. As Figuras 4.2(a) a 4.2(d), *esclerose pura sem crescente*; 4.2(e) a 4.2(h), glomérulos com *glomerulonefrite crescente*; 4.2(i) a 4.2(l) apresentam glomérulos com *nefropatia membranosa*; 4.2(m) a 4.2(p), *podocitopatia*.

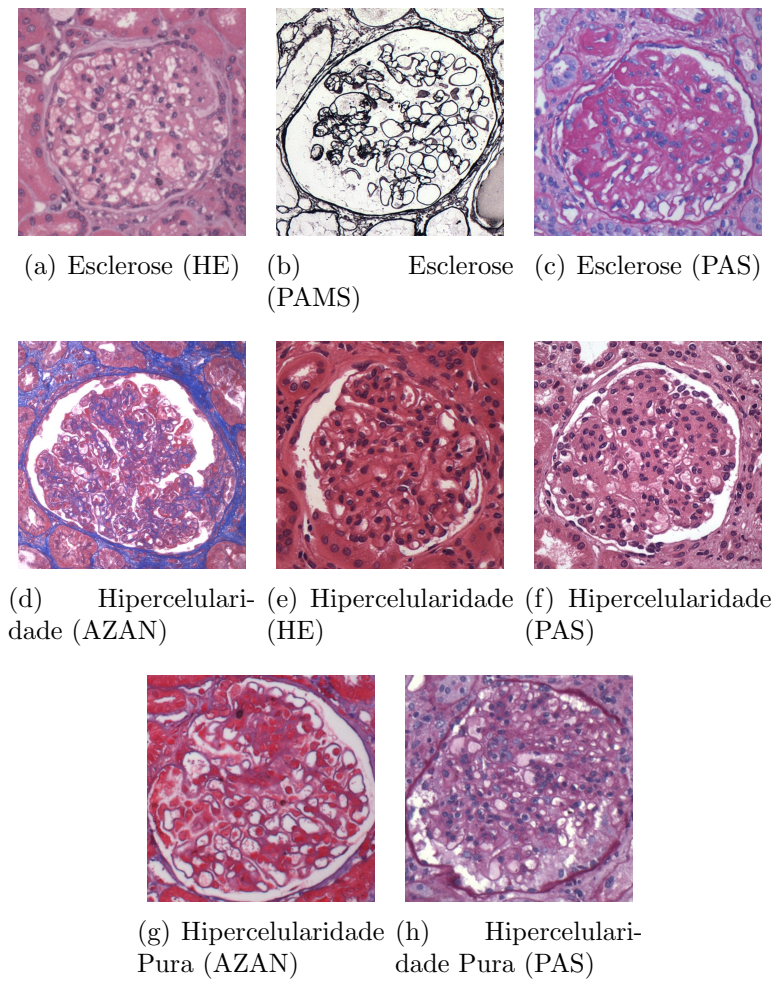


Figura 4.3: Imagens de glomérulos presentes no conjunto de dados. As Figuras 4.3(a) a 4.3(c), *esclerose*; 4.3(d) a 4.3(f), *hipercelularidade*; 4.3(g) a 4.3(h), *hipercelularidade pura sem crescente*.

4.2 O Conjunto de Dados Refinado

Como afirmado na seção anterior, o *dataset* original foi submetido a um processo de curadoria, com o intuito de reduzir possíveis ruídos introduzidos por amostras com características muito discrepantes das do restante do conjunto. Esse processo foi realizado de forma manual, com uma primeira etapa de verificação das imagens separando aquelas com a presença de algum ruído (distorções). Estas foram posteriormente agrupadas de acordo com o tipo de ruído apresentado, como explicado a seguir.

Para as finalidades deste estudo, a princípio, foi assumido que uma amostra considerada *adequada* deveria apresentar as seguintes características: (i) deveria apresentar um único glomérulo; (ii) o glomérulo representado deveria ocupar uma área entre

aproximadamente 50% e 75%¹ do quadro da imagem e, (iii) o glomérulo deveria aparecer de forma inteiriça na imagem, não sendo apenas parcialmente representado. Algumas das imagens com essas características podem ser visualizadas nas Figuras ?? e 4.2, enquanto a Figura 4.4 oferece um melhor entendimento dos limites aproximados para a ocupação desejada do glomérulo no quadro da imagem.

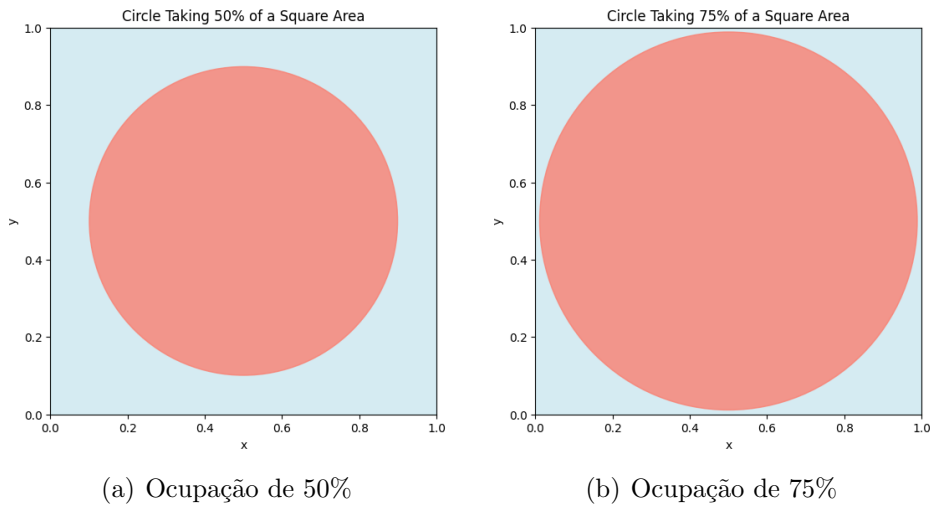


Figura 4.4: Ocupação aproximada desejável de um glomérulo dentro do quadro da imagem.

Com a análise do conjunto, verificou-se que 4.856, ou 41,6% das amostras apresentavam as características desejadas, enquanto 6818, ou 58,4% continham alguma característica considerada neste trabalho como *distorções*, que foram classificadas como: (i) distorção de ocupação, para amostras contendo glomérulos ocupando menos de 50,0% ou mais de 75,0% do quadro da imagem; (ii) distorção de representação, para os glomérulos parcialmente representados (cortados), nas imagens e, (iii) distorção de representação numérica, para amostras que apresentavam mais de um glomérulo. A Figura 4.5 apresenta alguns exemplos de amostras com distorções.

¹A escolha dos percentuais aqui citados foi feita de forma experimental, considerando o que foi observado na maioria das amostras do conjunto. Não há, neste trabalho, uma referência da literatura acadêmica para a escolha desses valores.

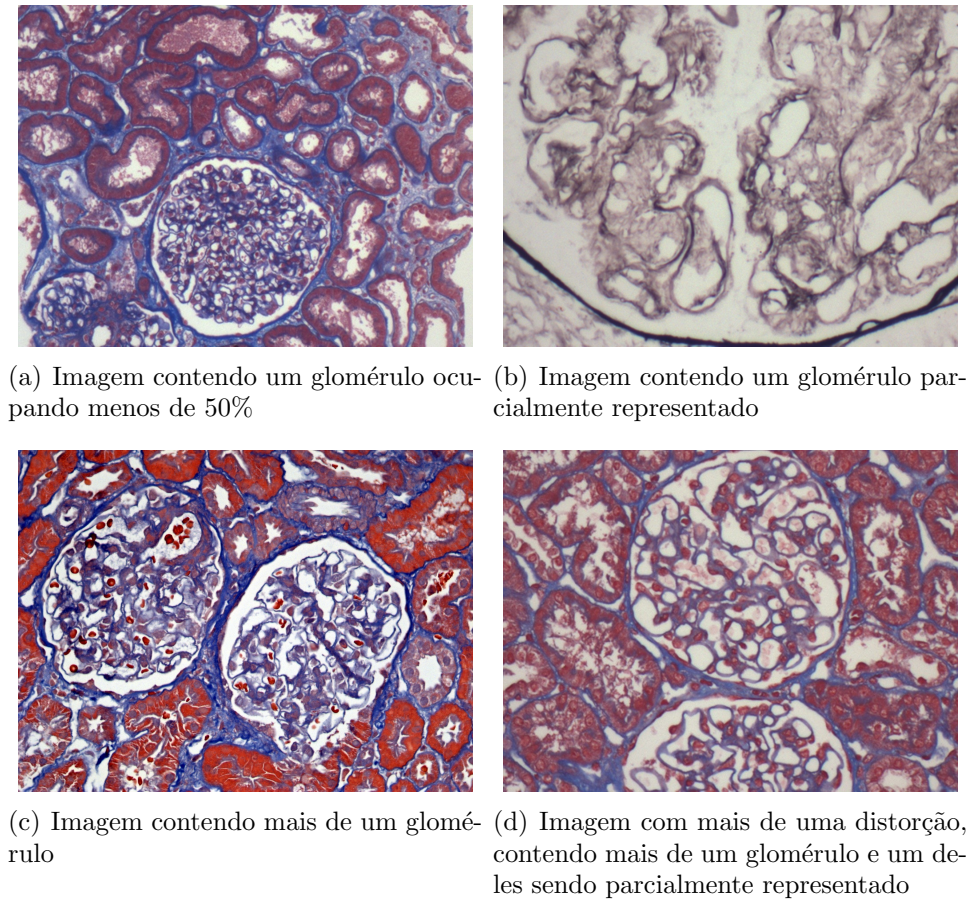


Figura 4.5: Distorções encontradas em amostras do conjunto.

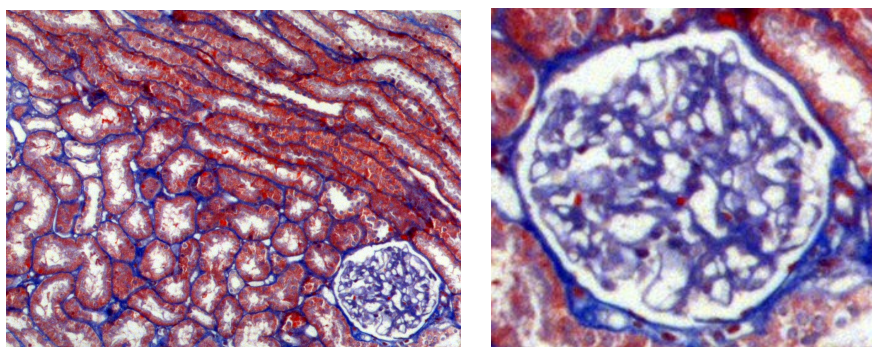
A distribuição das distorções encontradas é apresentada na Tabela 4.3, de forma segmentada para os corantes HE e PAS. A análise ainda revelou a existência de 62 amostras duplicadas, que foram removidas do conjunto. Salienta-se também o fato de que algumas amostras apresentaram mais de uma distorção.

Tabela 4.3: Distorções encontradas no conjunto.

Distorção	HE	PAS	Outros	Total
-	2.599	1.510	747	4.856
Ocupação	1.488	857	569	2.914
Representação Parcial	1.250	875	917	3.042
Representação Numérica	496	226	181	903

Após a identificação das amostras com distorções, foram adotadas medidas para a redução dos ruídos presentes, incluindo a edição das imagens sempre que possível. Nos casos em que a correção não era viável, optou-se pelo descarte das amostras

comprometidas. Para as amostras com distorção de ocupação, foi realizado o enquadramento do glomérulo de forma que este ficasse dentro dos limites estipulados. A Figura 4.6 ilustra um exemplo de imagem tratada para correção da distorção de ocupação. Todas as imagens cuja área ocupada ultrapassava o limite de 75,0% continham glomérulos parcialmente representados e foram assim classificadas.

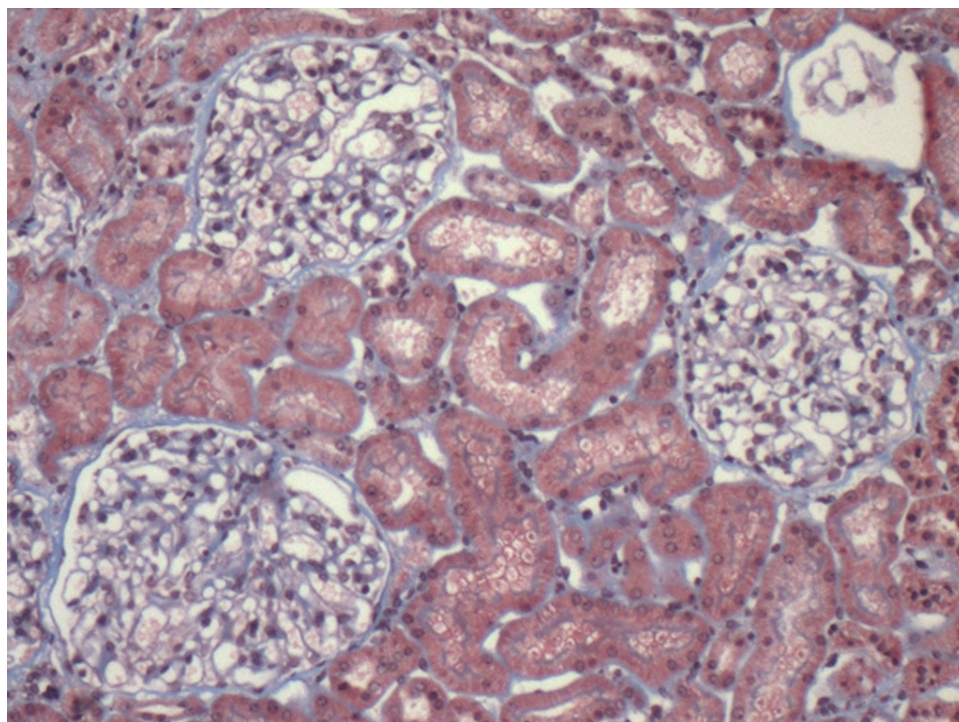


(a) Imagem não tratada apresentando um glomérulo com distorção de ocupação

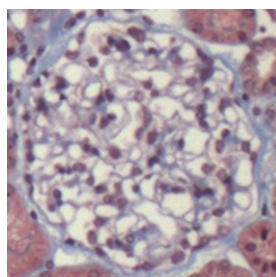
(b) Imagem corrigida

Figura 4.6: Tratamento de distorção de ocupação.

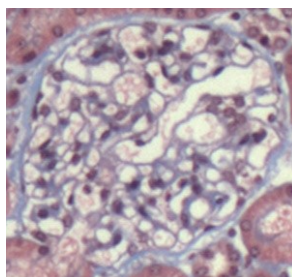
Para as amostras que continham mais de um glomérulo, foram realizadas edições com o objetivo de segmentá-las em múltiplas amostras individuais, descartando-se a imagem original após a separação. A Figura 4.7 ilustra o resultado desse procedimento.



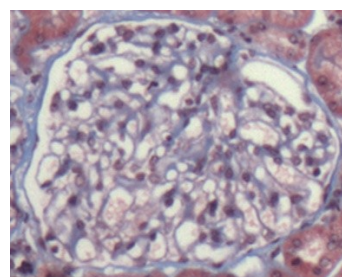
(a) Imagem original, contendo três glomérulos



(b) Amostra resultante nº 1



(c) Amostra resultante nº 2



(d) Amostra resultante nº 3

Figura 4.7: Tratamento para amostras com mais de um glomérulo.

As amostras contendo glomérulos parcialmente representados foram mantidas no conjunto, uma vez que constituem uma parcela significativa dos dados, estando especialmente presentes em classes já severamente desbalanceadas e com número reduzido de amostras, como amiloidose e hiper celularidade pura. A exclusão dessas imagens poderia acentuar ainda mais o desbalanceamento, comprometendo a viabilidade da construção de conjuntos com validação cruzada para o treinamento dos modelos. A Tabela 4.4 traz a distribuição de amostras com essa distorção, separada por classe.

Tabela 4.4: Percentual de imagens contendo glomérulos parcialmente representados, separado por classe.

Classe	Percentual
Amiloidose	43,58%
Normal	15,18%
Esclerose pura sem crescente	35,43%
Hipercelularidade	33,85%
Hipercelularidade pura sem crescente	52,23%
Glomerulonefrite crescente	12,23%
Nefropatia membranosa	29,76%
Esclerose	8,27%
Podocitopatia	24,36%

Ao final do processo de curadoria obteve-se um novo conjunto, composto pelas imagens editadas, descartando-se as amostras ruidosas, como explicado nesta seção. O novo conjunto é composto de 12.495 amostras e é descrito na Tabela 4.5.

Tabela 4.5: Distribuição do conjunto de dados curado por lesão e corante.

Classe	AZAN	HE	PAMS	PAS	Total
Amiloidose	31	145	96	102	374
Normal	259	1.717	366	580	2.922
Esclerose Pura	234	672	104	472	1.482
Hipercelularidade	263	2.069	0	1.067	3.399
Hipercelularidade Pura	60	0	0	164	224
Glomerulonefrite Crescente	125	521	183	380	1.209
Nefropatia Membranosa	148	851	327	431	1.757
Esclerose	0	279	122	219	620
Podocitopatia	90	65	108	245	508
Total	1.210	6.319	1.306	3.660	12.495

Tabela 4.6: Distribuição do conjunto de dados curado, por lesão e corante em valores percentuais.

Classe	AZAN	HE	PAMS	PAS	Total
Amiloidose	0,24%	1,16%	0,76%	0,81%	2,97%
Normal	2,07%	13,74%	2,93%	4,64%	22,39%
Esclerose Pura	1,87%	5,38%	0,83%	3,77%	11,86%
Hipercelularidade	2,10%	16,56%	-	8,54%	27,21%
Hipercelularidade Pura	0,48%	-	-	1,31%	1,79%
Glomerulonefrite Crescente	1%	4,17%	1,49%	3,09%	9,7%
Nefropatia Membranosa	1,18%	6,81%	2,62%	3,45%	14,1%
Esclerose	-	2,23%	0,98%	1,75%	4,97%
Podocitopatia	0,72%	0,52%	0,86%	1,96%	4,07%
Total	9,66%	50,57%	10,44%	29,27%	100%

É possível ver na Tabela 4.6, que não houve alteração na quantidade de amostras para as classes amiloidose, esclerose pura e hipercelularidade pura, o que contribuiu para um aumento na taxa de desbalanceamento para essas classes, já que, em relação ao conjunto original, foram introduzidas 821 novas amostras.

4.3 Ambiente Computacional

Os experimentos descritos neste trabalho foram executados sobre uma máquina HGX Next, composta por 8 GPUs NVIDIA A100 e 40GB de memória dedicada por GPU, processador AMD EPYC 7742 com 64 núcleos de 2.25GHz, 1TB de memória RAM e 20TB de armazenamento secundário. O código foi desenvolvido utilizando-se a linguagem Python versão 3.9.5, utilizando-se as bibliotecas PyTorch 2.4.0, Torchvision 0.19.0, NVIDIA Cuda 11.8.87 e Scikit Learn 1.5.1.

Para o treinamento de todos os modelos, adotou-se a *entropia cruzada* (*cross-entropy*) como função de custo e a *descida do gradiente estocástico* (*stochastic gradient descent* – SGD) como método de otimização, ambas amplamente empregadas em tarefas de classificação supervisionada (Goodfellow et al., 2016). A taxa de aprendizado inicial foi fixada em 0,01, valor comumente utilizado em conjunto com redes neurais convolucionais, especialmente ao se utilizar otimizadores como Adam e SGD (Bengio, 2012).

Para favorecer a convergência dos modelos, ajustes dinâmicos foram realizados por meio de um escalonador (*scheduler*) durante os treinamentos. O uso de mecanismos de escalonamento dinâmico é recomendado para evitar que a otimização fique presa em mínimos locais, favorecendo a convergência a soluções mais generalizáveis (Paszke et al., 2019). Os ajustes são feitos através da redução da taxa de aprendizado por um fator γ (*gamma*), dado um determinado número de épocas (*step*). Para os experimentos realizados neste trabalho, foi utilizado um *step* igual a 32, por ser o tamanho dos lotes de treinamento e um fator γ de 0,1, aplicando-se ainda um *piso*

de 10^{-6} para evitar o desaparecimento do gradiente (*gradient vanishing*)², valores comumente utilizados na prática e em trabalhos acadêmicos (Paszke et al., 2019; He et al., 2015), (Bengio, 2012). Como mencionado, o tamanho dos lotes (*batches*) foi mantido fixo em 32, valor que oferece um equilíbrio entre estabilidade no treinamento e uso eficiente dos recursos computacionais (Masters e Luschi, 2018; Bengio, 2012).

4.4 Organização dos Conjuntos de Dados para os Experimentos

Para a realização das tarefas descritas neste texto, foram construídos conjuntos de dados a partir do conjunto original, descrito nas Tabelas 4.1 e 4.5. Esses conjuntos foram nomeados de acordo com suas características para serem melhor referenciados ao longo deste texto e são listados na Tabela A.1, no Apêndice A.

Todos os conjuntos foram construídos para tarefas binárias, com cada conjunto adotando uma das classes como sendo a positiva e agrupando todas as demais em uma única classe negativa, seguindo a estratégia de decomposição *um contra todos* (*one versus all* - OVA) (Fernandez et al., 2018). Para facilidade de entendimento, os conjuntos que utilizam uma determinada classe como positiva, foram nomeados de acordo com a classe de interesse. Conjuntos *baseline* apresentam amostras de todas as classes e todos os corantes, com superamostragem aleatória (ROS); conjuntos restritos a uma um corante, apresentam o nome do corante associado; conjuntos *curated* são aqueles originados a partir dos dados refinados no processo de curadoria.

Em todos os experimentos, os conjuntos de dados foram divididos, aplicando-se uma partição *K-fold* para validação cruzada, na qual, a partir do conjunto de origem, são produzidos *K* conjuntos (*folds*). Utilizando $K = 5$, a configuração para cada *fold* ficou definida com 80% dos dados utilizados para treino e 20% (disjunto) reservados para validação. O valor para *K* foi definido de forma a não reduzir demasiadamente a quantidade de amostras de validação por *fold* para as classes menos representadas, o que poderia comprometer a avaliação dos modelos, já que mesmo poucos erros de classificação poderiam representar variações significativas.

O treinamento dos modelos considerados como *baseline*, foi realizado sobre os conjuntos nomeados como *Baseline-[nome da classe]*, nos quais nenhuma técnica de reamostragem foi aplicada. Esse conjunto de experimentos estabelece a referência de comparação para as outras técnicas aqui empregadas. Também foram realizados experimentos com conjuntos sobre os quais a superamostragem aleatória foi aplicada, balanceando a distribuição entre amostras positivas e negativas na proporção 1:1. A partir do conjunto original, ainda foram produzidos *datasets* restritos à presença de amostras tingidas com apenas um corante. Embora haja quatro corantes

²O desaparecimento do gradiente ocorre quando os valores dos gradientes calculados durante a retropropagação (*back propagation*) tornam-se progressivamente menores nas camadas mais profundas da rede, dificultando ou impedindo a atualização efetiva dos pesos (Goodfellow et al., 2016).

presentes no conjunto descrito neste documento, foram selecionados apenas as colorações HE e PAS. O corante HE foi selecionado por ser referenciado na literatura com a mais popular coloração histológica (Montanari, 2016), o que pode ser observado no conjunto aqui estudado, em que amostras com HE representam cerca de 50% de todos os dados. Já o corante PAS foi selecionado por ser o segundo mais bem representado, além de ser o único corante a apresentar amostras para todas as classes.

Procedimento semelhante foi realizado sobre conjunto refinado pelo processo de curadoria. Foram produzidos conjuntos com todas as classes e corantes, sem o uso de rebalanceamento; conjuntos sobre os quais se aplicou a superamostragem aleatória; conjuntos restritos à presença de amostras tingidas com um único corante, também restritos a HE e PAS.

4.5 Modelos Convencionais Usando CNN

A construção dos modelos utilizou as arquiteturas Kolmogorov-Arnold ConvNeXt Network (KCN) (Cheon, 2024), EfficientNet-B0 (Tan e Le, 2019), EfficientNet-B4 (Tan e Le, 2019), Inception-v3 (Szegedy et al., 2015), ResNet-18 (He et al., 2015), ResNet-50 (He et al., 2015) e VGG-16 (Simonyan e Zisserman, 2014), todas pré-treinadas com o conjunto ImageNet, que é composto por 1.000 classes, com 1.000 amostras cada, totalizando 1.000.000 de imagens. Durante o treinamento, foi realizado um processo de *fine tuning*, em que todas as camadas da rede de *backbone* foram *descongeladas*, permitindo a atualização de seus pesos, habilitando os modelos a aprender as características dos dados de entrada, partindo de um conhecimento prévio, ao invés da inicialização randômica de seus pesos.

Para cada arquitetura, foi realizada uma iteração de 100 épocas sobre cada um dos cinco *folds*, resultando na produção de um modelo por *fold*. Adotou-se o critério de parada antecipada, interrompendo o treinamento após 10 épocas consecutivas sem melhora no desempenho. Foi também testado o uso de parada antecipada com 20 épocas, porém, esta configuração não trouxe ganhos de desempenho. Todos os modelos foram treinados com a aplicação de aprendizado sensível ao custo (*cost-sensitive learning*), ponderando-se a função de perda com pesos inversamente proporcionais à representatividade de cada classe, como apresentado nas Equações 4.1 e 4.2, nas quais, o número de amostras das classes positiva e negativa é dividido pelo total de amostras no conjunto. A descrição completa dos pesos utilizados em cada configuração de treinamento encontra-se no Apêndice B, bem como as respectivas taxas de desbalanceamento (IR). Os resultados foram obtidos a partir da média e do desvio padrão do F1-score, calculados com base no desempenho dos modelos treinados sobre os cinco *folds*.

$$positive_class_weight = \frac{numero_de_amostras_majoritarias}{numero_total_de_amostras} \quad (4.1)$$

$$negative_class_weight = \frac{numero_de_amostras_minoritarias}{numero_total_de_amostras} \quad (4.2)$$

4.6 Modelos Few-shot Learning

Os experimentos com FSL foram realizados com os conjuntos *Baseline-[nome da classe]*, *HE-[nome da classe]*, *PAS-[nome da classe]*, *Curated-[nome da classe]*, *Curated-HE-[nome da classe]* e *Curated-PAS-[nome da classe]*. Sobre esses conjuntos, não foi aplicada a superamostragem aleatória.

No desenvolvimento dos modelos baseados em *Few-Shot Learning* (FSL), adotou-se uma abordagem baseada em métrica, utilizando redes prototípicas, pela sua eficiência computacional e pela vantagem de permitir a criação de uma representação em nível de classe, através dos protótipos, diferente das *matching networks* e redes siamesas, conforme discutido no Capítulo 2. Para as classes com taxa de desbalanceamento igual ou inferior a aproximadamente 1 : 10, os treinamentos foram conduzidos por meio de episódios do tipo 2-way 200-shot, aplicados sobre a partição de treinamento de cada *fold*, com *query sets* compostos por 50 amostras, buscando-se manter, sempre que possível, a proporção de 80% : 20% entre suporte e consulta nos episódios. Outras configurações foram testadas, como 100-shot com 25-query, 400-shot com 100-query e 800-shot com 200-query; contudo, essas variações não resultaram em ganhos e, em alguns casos, causaram decréscimos de desempenho, como foi observado para a classe *hipercelularidade*.

Para classes com desbalanceamento mais acentuado (taxas superiores a 1 : 10), os protótipos foram construídos com até 80% das amostras positivas disponíveis na partição de treinamento, mantendo-se a proporção 80% : 20% entre suporte e consulta nos episódios. No entanto, o maior desempenho foi obtido ao se utilizar 25% da porção de treinamento como *support set*, como é demonstrado no Capítulo 5 para as classes *amiloidose*, *esclerose* e *podocitopatia*.

A definição da quantidade de amostras por episódio levou em consideração a disponibilidade de dados por classe, com o objetivo de manter os episódios de treinamento balanceados e evitar o viés em favor da classe majoritária, comportamento semelhante ao adotado em estratégias de subamostragem. Para isso, limitou-se a quantidade de amostras da classe majoritária na construção dos protótipos, o que também pode ter contribuído para a redução de potenciais ruídos, decorrentes de um número excessivo de exemplos. Como discutido anteriormente, o aumento no tamanho do *support set* não resultou em ganho de desempenho, o que sugere que o uso de um número reduzido de amostras pode favorecer a extração de características mais representativas para cada classe.

Após o treinamento, os modelos foram avaliados sobre a partição de validação correspondente a cada *fold*. A Figura 4.8 ilustra o processo de treinamento episódico empregado neste estudo. Cada *fold* foi treinado por 100 épocas, com aplicação de

parada antecipada (*early stopping*) após 10 épocas consecutivas sem melhoria na função de custo. Em cada época, foram executados 100 episódios. Configurações alternativas, como o uso de 200 épocas e parada antecipada após 20 épocas, também foram avaliadas, mas não resultaram em aumento no desempenho. Da mesma forma, o número de episódios por época foi testado com valores de 50 e 200, sem impacto positivo sobre os resultados finais.

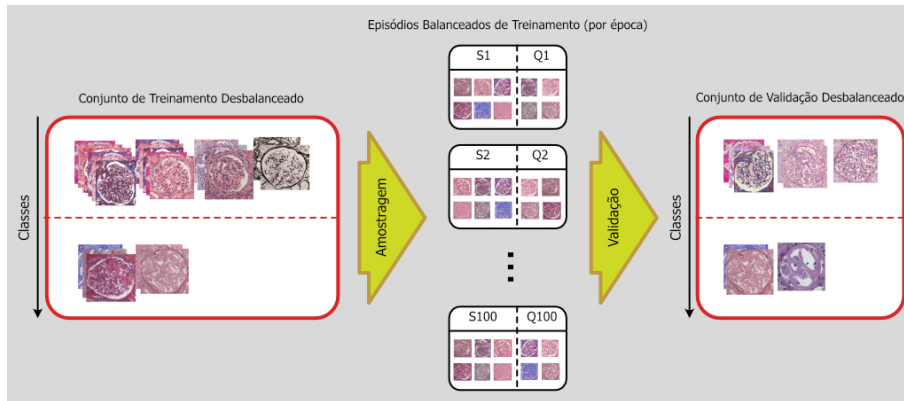


Figura 4.8: Treinamento episódico balanceado sobre um conjunto desbalanceado. Para cada *fold*, foram conduzidas até 100 épocas, com 100 episódios por época. A validação final foi realizada sobre dados desbalanceados.

Para a extração de características, foram utilizadas as mesmas arquiteturas de *backbone* empregadas nos experimentos com métodos convencionais, aproveitando-se os parâmetros dos modelos salvos para cada *fold*. Por exemplo, tomando-se o modelo produzido com a arquitetura Inception-v3 sobre o *fold 1* de um determinado conjunto utilizando métodos comuns, os parâmetros desse modelo foram empregados no extrator de características para o treinamento do modelo FSL sobre o mesmo *fold*, usando a mesma arquitetura. Dessa forma, para cada *fold* treinado com FSL, utilizaram-se os pesos correspondentes obtidos com os métodos comuns, assegurando-se que não houvesse vazamento de dados entre as etapas de treinamento e validação. A motivação para o uso de tal técnica reside na ideia de que os parâmetros obtidos sobre dados de mesmo domínio ofereceriam um ponto de partida mais adequado do que o oferecido por aqueles obtidos sobre o conjunto ImageNet.

Como função de cálculo de distância entre os vetores de características, foi utilizada a similaridade do cosseno, tendo a distância euclidiana sido avaliada, mas descartada devido ao desempenho inferior dos classificadores ao utilizá-la, o que será demonstrado no Capítulo 5. Uma possível razão para este comportamento pode estar na própria natureza da similaridade do cosseno, na qual, o que é considerado é a direção apontada pelos vetores de características, ao invés de sua magnitude ou distância absoluta, o que faz com que essa função seja mais robusta contra variações de cor ou escala.

Um ponto importante deste estudo é que o diferencia dos trabalhos citados no Capítulo 3 é o fato de que, apesar de se ter empregado o treinamento episódico, típico em tarefas de FSL, todas as avaliações finais dos modelos foram realizadas sobre dados desbalanceados, ao invés de se utilizar apenas a validação pelo conjunto de consulta, dentro de cada episódio. Esse processo de avaliação dos modelos foi escolhido por oferecer um aferidor justo para comparação da técnica com o que foi realizado com os métodos comuns.

4.7 Modelos Ensemble-based

Além dos recursos utilizados nas etapas de pré-processamento como o uso de superamostragem e *data augmentation*, e do uso de *cost sensitive learning*, também foi adotada a utilização de modelos em conjunto (*ensemble-based*), como estratégia para obtenção de ganho no desempenho dos modelos utilizados. Como mencionado no Capítulo 2, modelos *ensemble-based* são abordagens comuns utilizadas em cenários de desbalanceamento, geralmente obtendo resultados superiores àqueles alcançados por um único classificador.

Para cada classe estudada neste trabalho, foi construído um classificador *ensemble-based*, utilizando os parâmetros dos três modelos que obtiveram os maiores valores de F1-score no conjunto de dados que apresentou o maior desempenho global para a respectiva classe de interesse, através de um processo de classificação por votação simples.

4.8 Análise da Complexidade dos Dados

Para análise dos conjuntos utilizados neste trabalho foram selecionados quatro métodos: (i) análise de componentes principais (PCA), para uma visualização da distribuição das amostras no espaço de características; (ii) fração de pontos de borda (N1), servindo como uma medida de sobreposição entre classes, utilizando um classificador de vizinhos mais próximos (*K-nearest neighbors* - KNN); (iii) silhouette, para análise da relação entre compactação e sobreposição das classes e; (iii) Métrica de separabilidade linear, utilizando um classificador baseado em SVM, observando-se o seu desempenho com respeito à métrica F1-score. Em todos os métodos foram utilizados os vetores de características extraídos com arquitetura a ter alcançado o maior desempenho para F1-score durante o treinamento, para cada conjunto e os valores foram computados sobre os cinco *folds* de cada conjunto em questão, calculando-se média e desvio padrão.

4.9 Comparação Estatística dos Modelos

Considerando que este trabalho comparou o desempenho de modelos construídos com diferentes configurações e dados, tornou-se necessária a adoção de um método

para medir a significância estatística dos diferentes resultados obtidos. O desempenho dos modelos foi medido com base na média e desvio padrão alcançados sobre os cinco *folds* utilizados em seu treinamento.

Para a avaliação da significância estatística entre os desempenhos, foi adotado o teste de postos com sinais de Wilcoxon (*Wilcoxon signed-rank test*), conforme os fundamentos previamente discutidos no Capítulo 2. O teste foi conduzido para a avaliação de dois cenários: (i) comparação dos resultados obtidos pelos modelos convencionais e aqueles construídos com redes prototípicas; (ii) comparação entre os resultados obtidos sobre conjuntos multicolorante e aqueles segmentados por um único corante.

Em cada comparação realizada, considerou-se como hipótese nula (H_0) a inexistência de diferença estatisticamente significativa entre os modelos avaliados, rejeitando-se essa hipótese nos casos em que o valor- p obtido foi inferior a 0,10. Embora valores mais conservadores, como 0,05, sejam comumente utilizados na literatura, a escolha de um limiar mais flexível se justifica pela baixa cardinalidade dos dados avaliados (McDonald, 2014) — uma vez que as comparações entre modelos são realizadas com base em seus desempenhos em apenas 5 *folds*. Dessa forma, a distribuição dos resultados não oferece dados suficientes para a adoção de um limiar mais restritivo.

Capítulo 5

Resultados e Discussão

Este capítulo apresenta os resultados obtidos por este trabalho. Os resultados apresentados aqui foram consolidados por classe, tomando-se as primeiras dez maiores taxas de F1-score obtidas nos experimentos. Uma relação mais completa de resultados, contendo as três maiores taxas atingidas por cada classe em cada conjunto¹, é apresentada no Apêndice C, enquanto o Apêndice D apresenta os gráficos da função de custo, curvas *precision-recall* e matrizes de confusão obtidos sobre os cinco *folds* para o modelo de maior desempenho. Salienta-se que este estudo adotou F1-score como sendo a métrica principal de aferição de desempenho dos modelos, tomando *precision* e *recall* como métricas secundárias.

Neste capítulo, dedicou-se uma seção para a análise de cada classe e a última seção traz uma discussão geral dos resultados. Cada seção é iniciada com uma descrição do conjunto a ter gerado o modelo com maior desempenho, com uma análise de sua complexidade com base nas métricas adotadas neste trabalho.

Com respeito à comparação estatística entre os resultados obtidos, esta foi realizada utilizando-se o teste de Wilcoxon, considerando-se a existência de significância estatística para os casos em que o valor-*p* obtido foi inferior a 0,10, pelas razões expostas no Capítulo 4.

5.1 Resultados e Análise para a Classe Amiloidose

Para a classe *amiloidose*, o maior desempenho foi alcançado sobre o conjunto *HE-amiloidose*, como é visto na Tabela 5.2. As métricas de complexidade obtidas para essa classe sugerem um conjunto com complexidade moderada. Como mencionado anteriormente, com base nos critérios adotados neste estudo, foi realizada uma análise de componentes principais (*Principal Component Analysis* – PCA), utilizando os

¹Neste trabalho, os conjuntos de dados são nomeados de acordo com a classe de interesse, conforme o contexto. Nas seções a seguir, a designação da classe foi omitida da nomenclatura dos conjuntos nos casos em que tal informação pôde ser inferida de maneira implícita.

vetores de características extraídos pela rede convolucional *ResNet-50* utilizando-se os pesos salvos durante o treinamento. O resultado é exibido na Figura 5.1, onde observa-se pontos da classe positiva dispersos no espaço de características. Também foram calculados os valores para *silhouette*, *N1* e *separabilidade linear*, exibidos na Tabela 5.1.

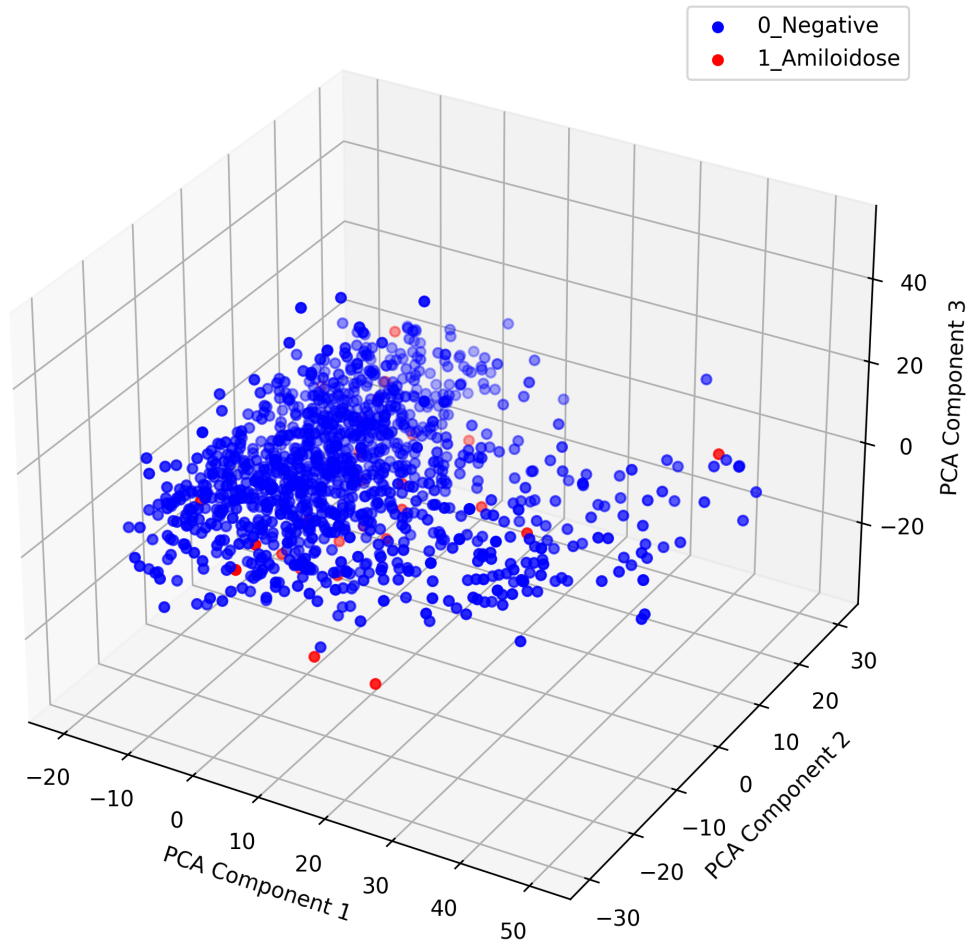


Figura 5.1: Análise de componentes principais para a classe *amiloidose*.

O valor obtido para *N1* sugere uma fronteira de decisão com um nível de sobreposição moderado. O índice obtido para *silhouette* indica que a *amiloidose* é uma classe com baixa compactação, com pontos de dados espalhados pelo espaço vetorial, favorecendo a ocorrência de áreas de sobreposição, o que é consistente com o observado na decomposição PCA. Já o índice obtido para *linearidade* sugere uma classe com separabilidade acentuada. Os valores observados apresentam consistência com o desempenho obtido para essa classe, apesar de sua elevada taxa de desbalanceamento.

Tabela 5.1: Medidas de complexidade para a classe *amiloide*.

Métrica	Valor
N1	0,330($\pm 0,076$)
Separabilidade Linear	0,782($\pm 0,039$)
Silhouette	0,228($\pm 0,036$)

Analisando-se os resultados obtidos para *amiloide*, na Tabela 5.2, é possível observar a predominância dos resultados obtidos sobre conjuntos segmentados pelo corante HE. Um resultado que pode ser justificado pela redução de sub-conceitos causada pela segmentação, além do fato de que o corante HE é aquele com maior número de amostras para *amiloide*. Um outro aspecto que merece atenção está no fato de que não existem amostras para *hipercelularidade pura sem crescente* para esse corante, o que também contribui para uma redução de sua complexidade. A Tabela 5.2 também mostra que o modelo *ensemble-based* permitiu um aumento de 1,13% de desempenho em relação ao maior resultado de um modelo individual sobre o mesmo conjunto.

Tabela 5.2: Taxas mais altas de F1-score alcançadas para *amiloide*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas ResNet-50, EfficientNet-B0 e Inception-v3, sobre o conjunto HE.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(FSL)	HE	90,0% $\pm$ 4,2%	97,9% $\pm$ 4,6%	93,7% $\pm$ 3,3%
ResNet-50 _(FSL)	HE	90,4% $\pm$ 4,0%	95,2% $\pm$ 5,2%	92,6% $\pm$ 3,1%
Inception-v3 _(CNN)	HE	87,1% $\pm$ 7,6%	95,2% $\pm$ 5,2%	90,9% $\pm$ 5,7%
EfficientNet-B0 _(FSL)	HE	84,0% $\pm$ 6,1%	95,2% $\pm$ 3,1%	89,1% $\pm$ 3,7%
Inception-v3 _(FSL)	HE	88,2% $\pm$ 10,9%	91,0% $\pm$ 7,9%	89,1% $\pm$ 7,2%
Inception-v3 _(CNN)	Curated-HE	83,4% $\pm$ 8,2%	93,8% $\pm$ 3,8%	88,2% $\pm$ 6,1%
ResNet-18 _(CNN)	HE	82,9% $\pm$ 8,5%	94,5% $\pm$ 7,2%	88,2% $\pm$ 7,1%
ResNet-50 _(FSL)	PAS	84,2% $\pm$ 6,7%	93,1% $\pm$ 5,6%	88,1% $\pm$ 2,7%
ResNet-50 _(FSL)	Curated-HE	86,2% $\pm$ 4,9%	89,0% $\pm$ 5,7%	87,4% $\pm$ 3,9%
Inception-v3 _(CNN)	Baseline	82,1% $\pm$ 2,4%	92,8% $\pm$ 1,2%	87,1% $\pm$ 1,2%
Inception-v3 _(FSL)	Curated-HE	88,6% $\pm$ 7,3%	86,2% $\pm$ 7,3%	87,1% $\pm$ 4,5%

De forma geral, os modelos treinados com conjuntos não curados apresentaram desempenho superior àquele obtido em conjuntos refinados pelo processo de curadoria. Esse fato é verificado tanto nos modelos baseados em CNN quanto nos classificadores construídos com o uso de redes prototípicas.

Os modelos baseados em FSL alcançaram as médias mais altas para F1-score, ressaltando-se novamente redução do desempenho nos experimentos realizados com os conjuntos curados. A redução de desempenho observada nos conjuntos curados pode estar relacionada ao aumento da taxa de desbalanceamento, uma vez que não houve acréscimo no número de amostras para a classe *amiloide*. Durante o refinamento, o conjunto de *amiloide* permaneceu praticamente inalterado, passando

apenas por pequenas edições em quatro imagens que apresentavam distorções de ocupação.

No que diz respeito à significância estatística, ressalta-se o fato de o teste de Wilcoxon ter evidenciado a superioridade dos modelos baseados em rede prototípicas, com um valor- p de aproximadamente 0,0625, como pode ser visto na Seção 5.10, na Tabela 5.21. outro aspecto observado nos resultados é o alto desvio padrão obtido por alguns modelos, chegando a valores acima dos 7%, indicando uma instabilidade no aprendizado destes frente à variabilidade dos dados, o que pode estar relacionado à sua complexidade.

Quanto à superioridade dos resultados para os experimentos restritos ao corante HE, observa-se que esta coloração é a que mais apresenta amostras para amiloidose. Outro fator relevante é a ausência da classe hiper celularidade pura, o que pode ter contribuído para uma redução da complexidade dos dados, reduzindo um potencial sub-conceito.

A análise de predições incorretas aponta que a maior parte dos falsos positivos recai sobre a classe *nefropatia membranosa*. A distribuição dos erros de classificação, considerando a média entre os 5 folds, é exibida na Figura 5.2, na qual, a legenda apresenta a representatividade de cada classe dentro do conjunto.

Média das Predições Incorretas por Classe (Amiloidose)

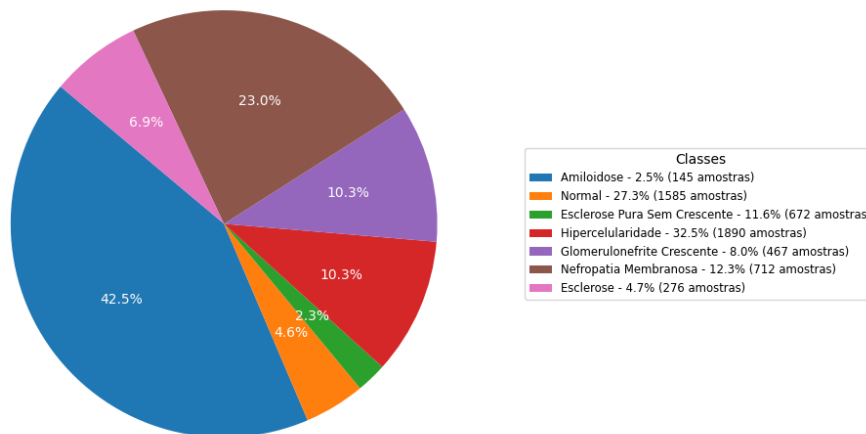


Figura 5.2: Distribuição das predições incorretas para *amiloidose*.

Com o objetivo de compreender melhor os erros de classificação, realizou-se uma investigação sobre as amostras responsáveis por falsos positivos e falsos negativos. A análise revelou que a maioria dessas amostras continha glomérulos parcialmente representados. Exemplos dessas imagens são apresentados nas Figuras 5.3 e 5.4, preservando-se a proporção original das amostras.

Diante do que foi observado, pode-se inferir que glomérulos parcialmente representados impõem um desafio adicional à tarefa de classificação, introduzindo potenciais

ruídos que podem comprometer a capacidade de generalização dos modelos. Contudo, destaca-se que aproximadamente 43% das amostras da classe *amiloidose* apresentam essa característica, o que inviabiliza sua exclusão sem causar uma redução drástica do conjunto de dados, já caracterizado por um alto grau de desbalanceamento, o segundo maior dentre as classes aqui estudadas.

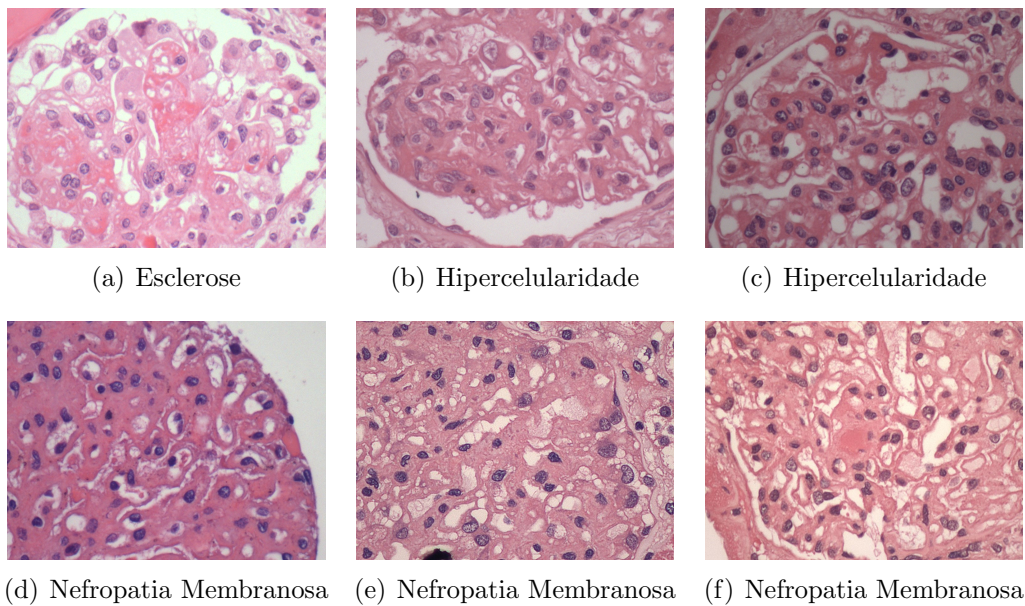


Figura 5.3: Amostras que produziram falsos positivos.

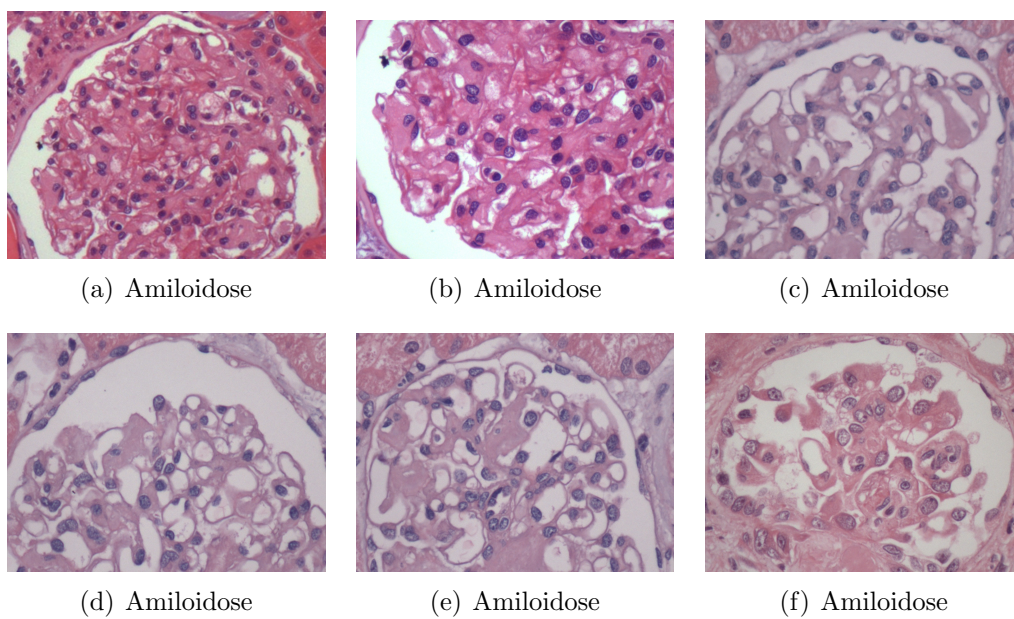


Figura 5.4: Amostras que produziram falsos negativos.

5.2 Resultados e Análise para a Classe Normal

A classe *normal* foi aquela a ter apresentado os maiores índices de separabilidade observados neste trabalho, o que é refletido no desempenho obtido nos experimentos. De forma geral, as métricas de complexidade indicam que essa classe apresenta uma alta separabilidade, fato que é evidenciado visualmente pela análise de componentes principais exibida na Figura 5.5, onde se observa a classe positiva quase totalmente separada da negativa, em um espaço de características tridimensional.

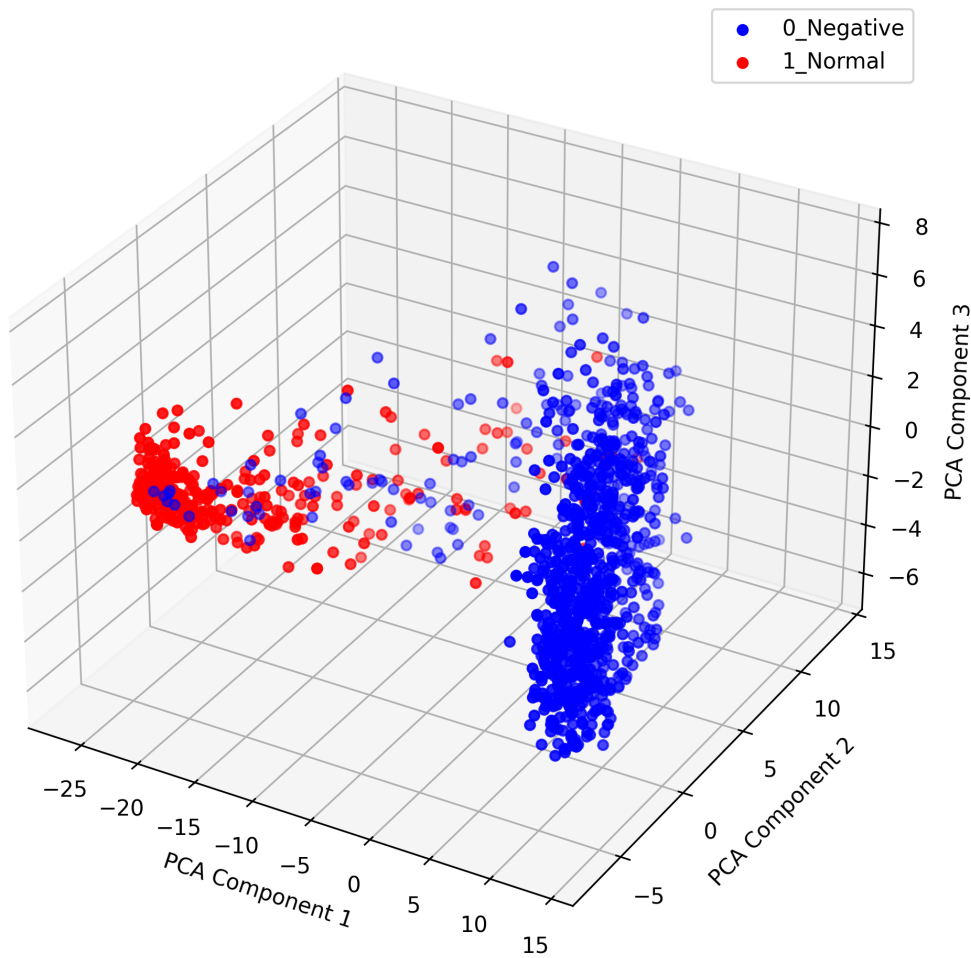


Figura 5.5: Análise de componentes principais para a classe *normal*.

Os resultados para as demais métricas de complexidade utilizadas neste trabalho também sugerem uma classe cujo conjunto apresenta relativa simplicidade em sua separação. A métrica para *separabilidade linear* obtida para normal, foi a maior de todas as nove classes estudadas, com um índice de linearidade de 86,9%, como

exibido na Tabela 5.3. O baixo $N1$ também sugere uma fronteira de decisão clara entre as classes positiva e negativa, enquanto o índice de *silhouette* sugere uma classe moderadamente compacta.

Tabela 5.3: Medidas de complexidade para a classe *normal*.

Métrica	Valor
$N1$	0,153($\pm 0,012$)
Separabilidade Linear	0,869($\pm 0,009$)
Silhouette	0,380($\pm 0,027$)

Para a classe *normal*, observou-se vantagem alcançada com os conjuntos refinados pelo processo de curadoria. Embora a taxa mais alta de F1-score tenha sido obtida sobre um conjunto não curado, a diferença para a segunda maior taxa é relativamente pequena e, além dessa observação, nota-se também predominância dos conjuntos refinados nos demais resultados em relação às suas contrapartes não curadas. Observa-se também aumento do desempenho para os experimentos realizados em conjuntos tratados com superamostragem ROS. A Tabela 5.4 exhibe os resultados para as dez taxas mais altas de F1-score e observa-se que o modelo *ensemble-based* obteve o maior desempenho.

Tabela 5.4: Taxas mais altas de F1-score alcançadas para *normal*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas Inception-v3, VGG-16 e ResNet-18, sobre o conjunto ROS-HE.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(CNN)	ROS-HE	97,6% $\pm$ 1,3%	97,7% $\pm$ 0,8%	97,7% $\pm$ 0,9%
Inception-v3 _(CNN)	ROS-HE	95,3% $\pm$ 1,3%	97,5% $\pm$ 1,3%	96,4% $\pm$ 0,6%
ResNet-18 _(CNN)	ROS-Curated-HE	94,7% $\pm$ 2,0%	98,0% $\pm$ 0,9%	96,3% $\pm$ 0,9%
VGG-16 _(CNN)	ROS-Curated-HE	93,6% $\pm$ 0,9%	98,8% $\pm$ 0,9%	96,1% $\pm$ 0,2%
Inception-v3 _(CNN)	Curated-HE	93,9% $\pm$ 1,7%	98,2% $\pm$ 0,5%	96,0% $\pm$ 0,8%
KCN _(CNN)	ROS-Curated-HE	93,4% $\pm$ 1,1%	98,7% $\pm$ 0,7%	96,0% $\pm$ 0,5%
VGG-16 _(CNN)	ROS-HE	94,0% $\pm$ 0,8%	97,7% $\pm$ 0,7%	95,8% $\pm$ 0,4%
VGG-16 _(CNN)	Curated-HE	93,2% $\pm$ 0,8%	98,5% $\pm$ 0,6%	95,8% $\pm$ 0,6%
ResNet-18 _(CNN)	Curated-HE	94,2% $\pm$ 1,7%	97,4% $\pm$ 1,0%	95,8% $\pm$ 0,9%
Inception-v3 _(FSL)	Curated-HE	93,0% $\pm$ 2,1%	98,7% $\pm$ 0,6%	95,7% $\pm$ 1,0%
ResNet-18 _(CNN)	ROS-HE	94,1% $\pm$ 2,4%	97,1% $\pm$ 1,8%	95,5% $\pm$ 0,8%

O desempenho mais alto foi obtido com o conjunto *ROS-HE-normal*, atingindo 96,41%($\pm 0,62\%$) de F1-score, seguido pelo conjunto *ROS-Curated-HE-normal*, com 96,29%($\pm 0,87\%$). Mais uma vez, os conjuntos restritos ao corante HE apresentaram desempenho superior em relação aos que utilizam amostras de múltiplos corantes, mesmo com a redução significativa no número de amostras, fator que, em tese, poderia comprometer a capacidade de generalização de redes convolucionais convencionais, como as empregadas neste trabalho. Ressalta-se novamente que, embora

a maior taxa de F1-score absoluta tenha sido alcançada com dados não curados, de forma geral, os conjuntos curados superaram suas respectivas contrapartes não refinadas na maioria dos cenários avaliados. Em todos os experimentos observou-se baixo desvio padrão, um possível indicativo da consistência do aprendizado dos modelos frente à variabilidade dos dados.

O uso de *Few-Shot Learning* (FSL) não apresentou ganhos de desempenho em comparação aos classificadores baseados em redes convolucionais (CNN), embora tenha alcançado taxas de F1-score superiores a 90% na maioria dos casos. Uma possível explicação para esse comportamento reside no fato de que o treinamento episódico, característico do FSL, realiza uma subamostragem das instâncias disponíveis para a construção dos protótipos, uma vez que nem todas as amostras são efetivamente utilizadas no conjunto de suporte em cada episódio. Tal característica pode prejudicar o desempenho em classes com maior representatividade, uma vez que a redução no uso das amostras disponíveis limita a capacidade do modelo de capturar a diversidade intra-classe. O teste de Wilcoxon evidenciou a superioridade dos modelos baseados em CNN, como mostrado na Tabela 5.21, na Subseção 5.10.1.

Os experimentos com FSL apresentaram desempenho ligeiramente abaixo ao alcançado com métodos comuns. De forma geral, observou-se vantagem nos modelos treinados com conjuntos curados, com o F1-score mais alto tendo sido obtido no conjunto *Curated-HE-normal* (95,72% ($\pm 0,96\%$)).

A análise das predições incorretas evidenciou que a *nefropatia membranosa* foi a classe mais presente entre os falsos positivos. Também observou-se a presença de muitas amostras com distorções de ocupação. A Figura 5.6 apresenta a distribuição de predições incorretas, enquanto as Figuras 5.7 e 5.8 mostram algumas amostras presentes entre os falsos positivos e falsos negativos respectivamente.

Média das Predições Incorretas por Classe (Normal)

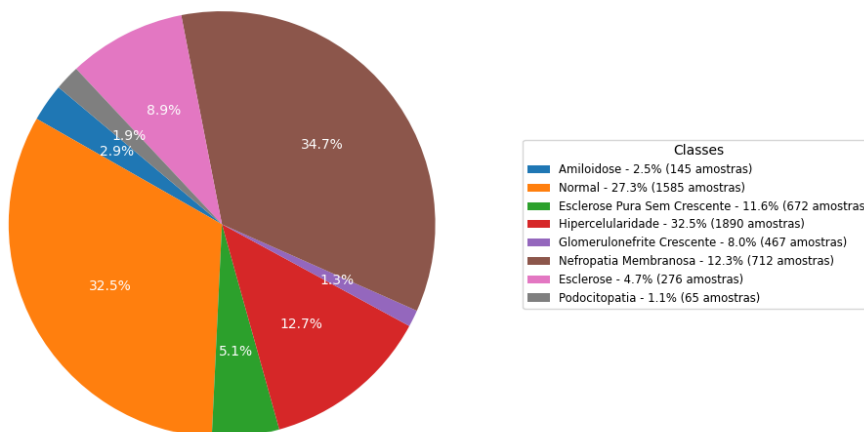


Figura 5.6: Distribuição das predições incorretas para *normal*.

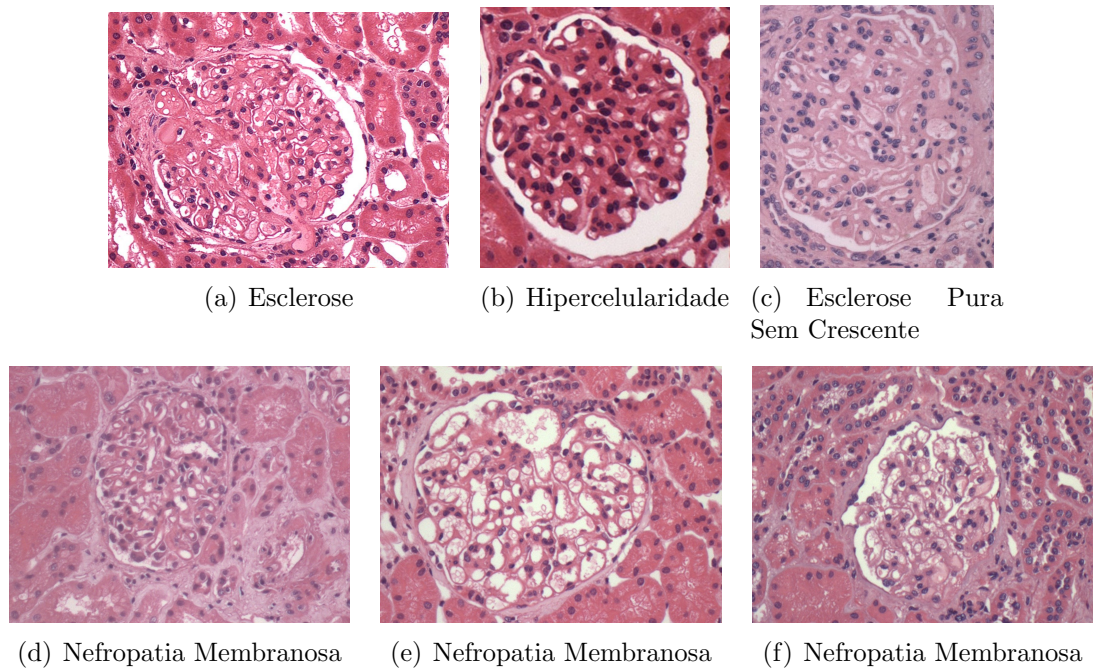


Figura 5.7: Amostras que produziram falsos positivos.

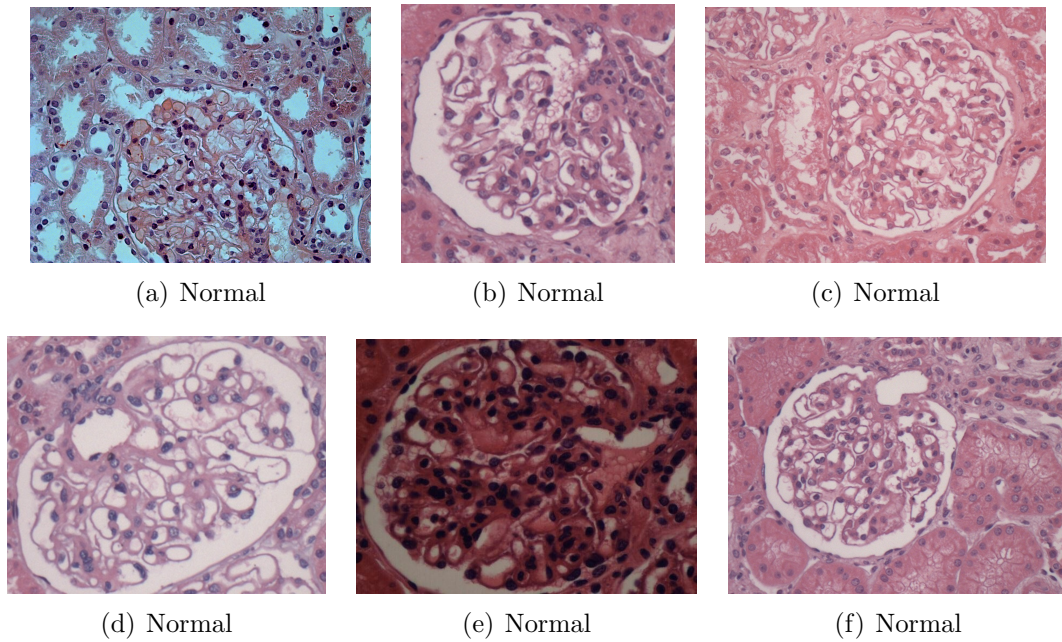


Figura 5.8: Amostras que produziram falsos negativos.

5.3 Resultados e Análise para a Classe Esclerose Pura sem Crescente

Os experimentos para a classe *esclerose pura sem crescente* apresentaram resultados com baixo F1-score. Considerando que essa classe apresenta uma taxa de desbalanceamento relativamente moderada (1:7), pode-se inferir que a complexidade do conjunto teve um papel importante para o desempenho observado. A decomposição PCA, aplicada ao conjunto *Curated-HE-esclerose-pura*, cujo resultado é apresentado na Figura 5.9, indica que as amostras de esclerose pura tendem a se concentrar junto às da classe negativa.

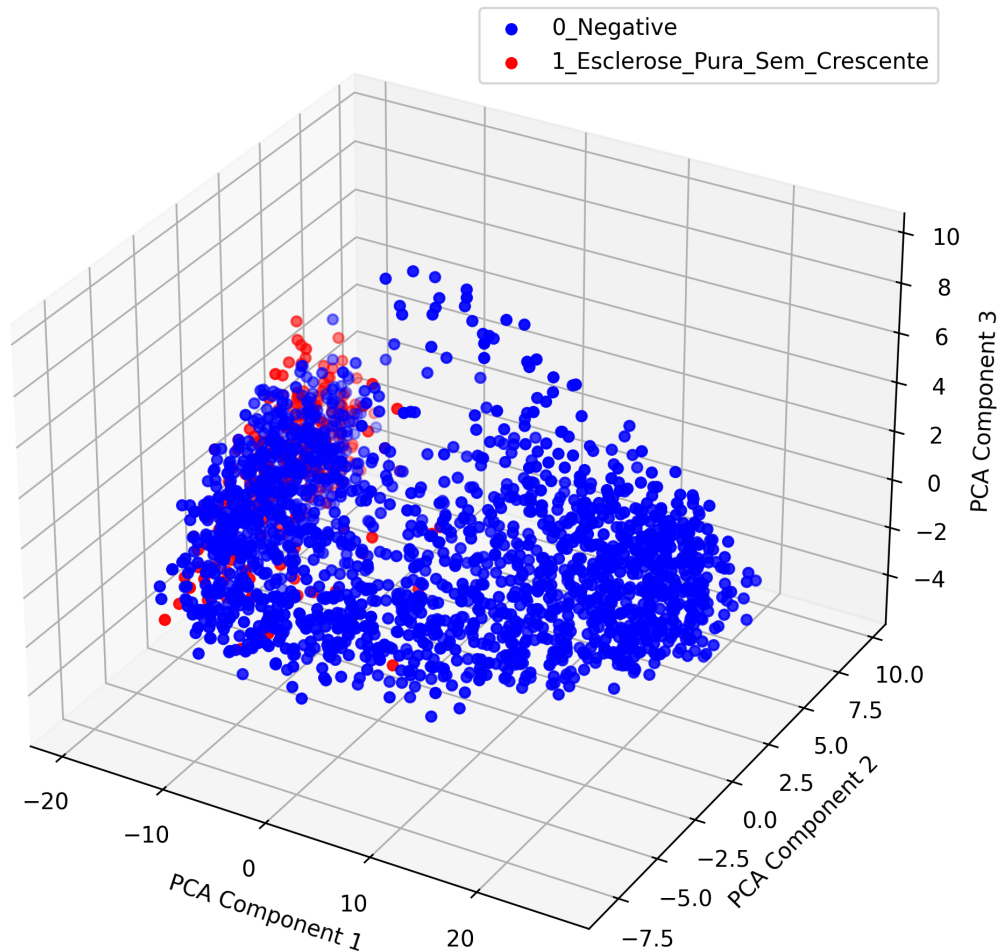


Figura 5.9: Análise de componentes principais para a classe *esclerose pura sem crescente*.

As métricas de complexidade obtidas para *esclerose pura sem crescente* evidenciam uma classe com baixa separabilidade. Destacam-se o alto valor obtido para $N1$, o

que indica uma fronteira de decisão com alto grau de sobreposição e, baixo valor alcançado para *silhouette*, sugerindo uma classe com baixa compactação. O índice obtido para linearidade indica uma classe não passível de ser separada linearmente. Os resultados são exibidos na Tabela 5.5.

Tabela 5.5: Medidas de complexidade para a classe *esclerose pura sem crescente*.

Métrica	Valor
N1	0,830($\pm 0,018$)
Separabilidade Linear	0,032($\pm 0,016$)
Silhouette	0,089($\pm 0,019$)

Como descrito na Tabela 5.6, o maior resultado foi alcançado com modelos baseados em CNN sobre dados não curados, sem restrição por corante e sem superamostragem (50,42%($\pm 1,31\%$)). Embora com uma taxa de desbalanceamento muito menor do que a apresentada para amiloidose, fica evidenciado que os conjuntos para esclerose pura sem crescente mostraram-se mais complexos, dificultando o aprendizado dos modelos. Os resultados de todos os experimentos com modelos individuais variaram entre 35,22%($\pm 2,97\%$) e 50,42%($\pm 1,31\%$), com o modelo *ensemble-based* alcançando 56,71%($\pm 4,79\%$).

Tabela 5.6: Taxas mais altas de F1-score alcançadas para *esclerose pura sem crescente*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas KCN, VGG-16 e EfficientNet-B0, sobre o conjunto *Baseline*.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble(CNN)	Baseline	40,5% $\pm$ 3,6%	94,4% $\pm$ 7,2%	56,7% $\pm$ 4,8%
KCN(CNN)	Baseline	35,5% $\pm$ 1,0%	86,9% $\pm$ 2,3%	50,4% $\pm$ 1,3%
VGG-16(CNN)	Baseline	36,5% $\pm$ 1,5%	80,3% $\pm$ 2,2%	50,2% $\pm$ 1,9%
KCN(CNN)	ROS	34,5% $\pm$ 0,2%	90,1% $\pm$ 1,9%	49,9% $\pm$ 0,4%
KCN(CNN)	HE	34,0% $\pm$ 1,2%	89,9% $\pm$ 3,2%	49,4% $\pm$ 1,4%
KCN(CNN)	ROS-HE	33,6% $\pm$ 1,2%	90,3% $\pm$ 4,2%	49,0% $\pm$ 1,5%
VGG-16(CNN)	HE	34,5% $\pm$ 0,9%	81,9% $\pm$ 2,8%	48,5% $\pm$ 0,6%
VGG-16(CNN)	ROS	35,5% $\pm$ 0,4%	75,0% $\pm$ 2,4%	48,2% $\pm$ 0,9%
EfficientNet-B0(CNN)	Baseline	34,2% $\pm$ 0,7%	79,1% $\pm$ 4,1%	47,7% $\pm$ 1,2%
VGG-16(CNN)	ROS-HE	34,1% $\pm$ 1,4%	76,9% $\pm$ 3,3%	47,2% $\pm$ 1,8%
EfficientNet-B0(CNN)	ROS	31,4% $\pm$ 0,8%	89,7% $\pm$ 2,3%	46,5% $\pm$ 1,1%

Os modelos construídos com *few-shot learning* (FSL) não apresentaram ganhos de desempenho em relação às redes convolucionais convencionais. Salienta-se que o teste de postos com sinais de Wilcoxon indicou significância estatística favorável aos modelos convencionais, evidenciando sua superioridade nos cenários avaliados. Outro aspecto relevante é o fato de que esta classe não teve incremento de novas amostras no processo de curadoria e apenas 7 imagens precisaram de ajustes quanto à ocupação dos glomérulos. Salienta-se também que, de forma geral, o desempenho sobre conjuntos curados foi inferior àquele obtido sobre os dados não refinados.

Observou-se ainda que, para todos os modelos, as taxas de precisão mantiveram-se consistentemente abaixo dos valores apresentados para *recall*, indicando uma alta taxa de falsos positivos. Para um melhor entendimento, foi realizada a análise das predições incorretas, o que revelou a predominância da classe *hipercelularidade* entre as ocorrências de falsos positivos.

A análise das predições incorretas apontou a *hipercelularidade* como a classe mais presente entre os falsos positivos, possivelmente, devido à sua alta representatividade no conjunto em evidência. A Figura 5.10 exibe a distribuição das predições incorretas para o conjunto *Baseline-esclerose-pura* e é possível observar que apenas 8,1% dos erros do modelo foram falsos negativos. As Figuras 5.11 e 5.12 apresentam algumas das amostras presentes nos erros de classificação.

Média das Predições Incorretas por Classe (Esclerose Pura Sem Crescente)

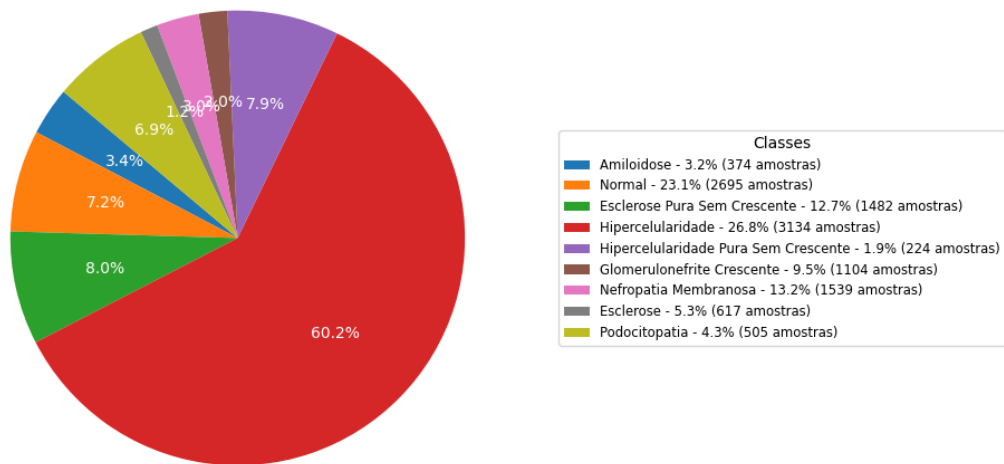


Figura 5.10: Distribuição das predições incorretas para *esclerose pura sem crescente*.

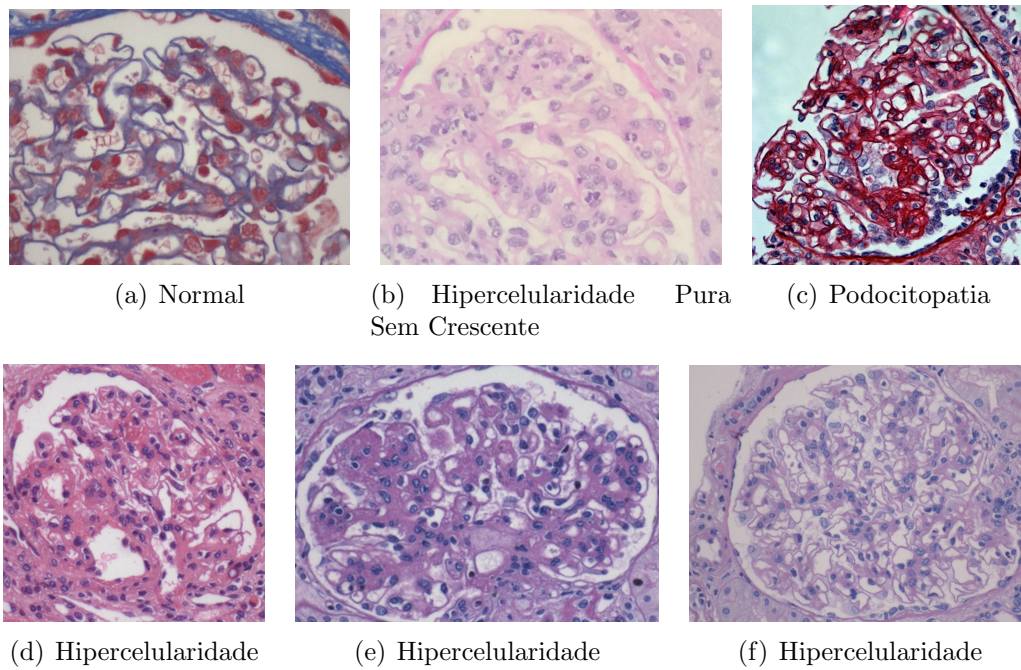


Figura 5.11: Amostras que produziram falsos positivos.

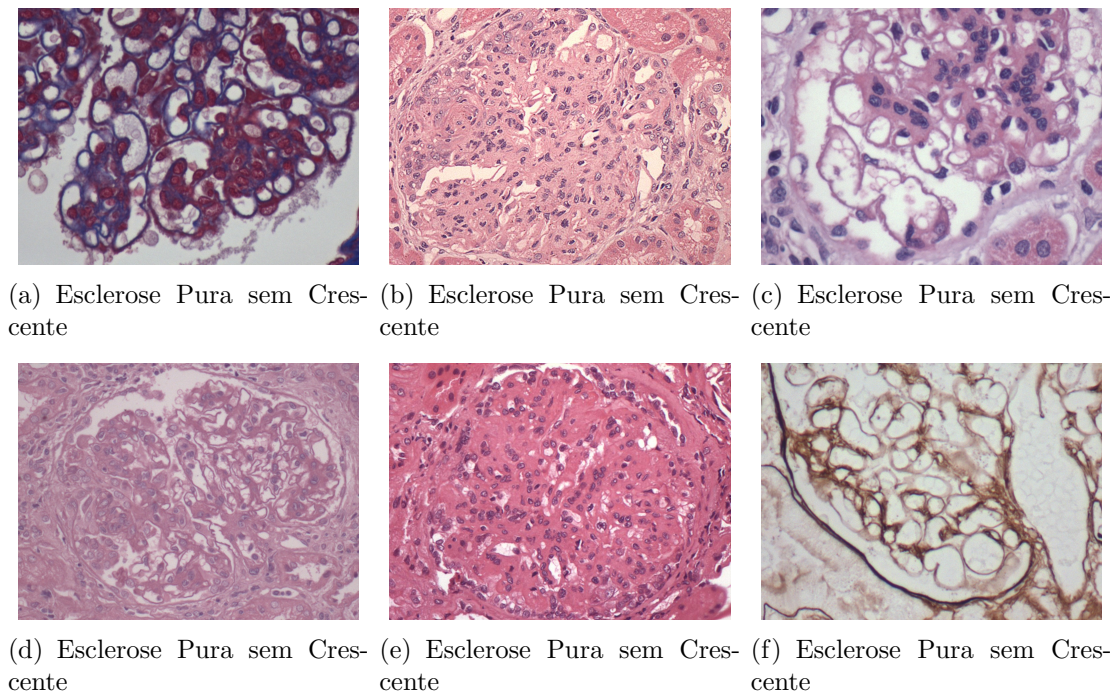


Figura 5.12: Amostras que produziram falsos negativos.

5.4 Resultados e Análise para a Classe Hiper celularidade

A *hipercelularidade* é a classe mais bem representada de todo o conjunto utilizado neste trabalho. No entanto, os modelos obtidos não apresentaram alto desempenho, com taxas de F1-score variando entre $67,97\%(\pm 1,20\%)$ a $78,91\%(\pm 1,51\%)$, com o valor mais alto tendo sido alcançado com um modelo *ensemble-based* construídos CNNs convencionais, sobre o conjunto *ROS-Curated-HE-hipercelularidade*.

Considerando a relativa baixa taxa de desbalanceamento e os resultados obtidos, é possível inferir que a complexidade do conjunto possa ter desempenhado um papel relevante para os resultados alcançados, indicando que, mesmo com baixo desbalanceamento, a classe em análise apresenta desafios para ser separada pelos classificadores. A análise de componentes principais apresentada na Figura 5.13, mostra que a classe positiva está apenas parcialmente separada, com diversas amostras situadas próximas ao espaço de características da classe negativa.

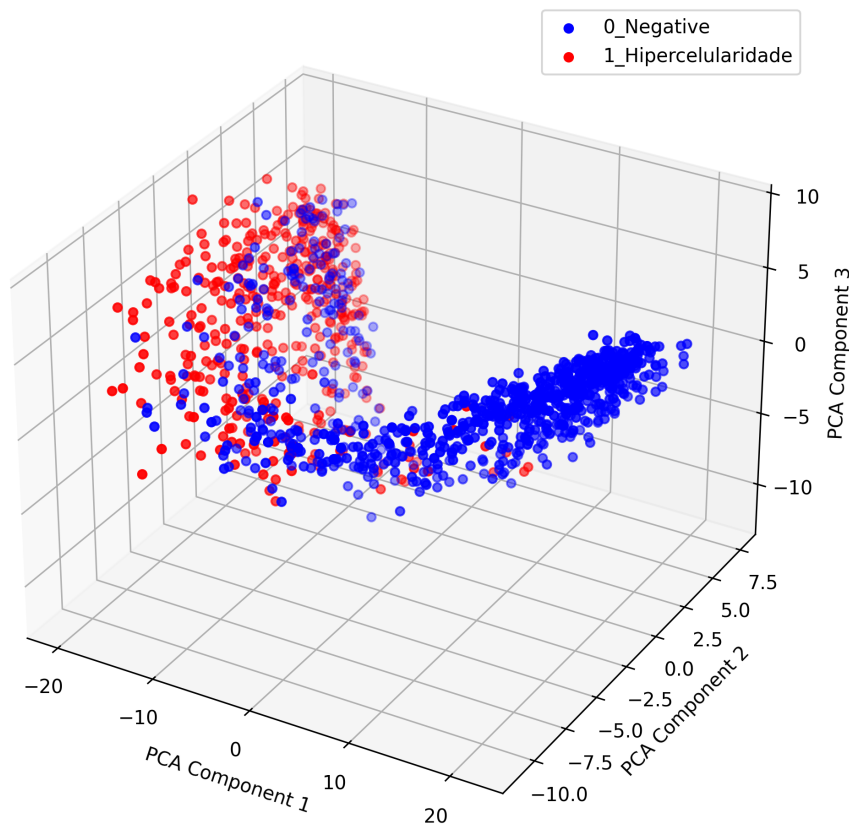


Figura 5.13: Análise de componentes principais para *hipercelularidade*.

As métricas de complexidade observadas sugerem um nível moderado de separabilidade. O valor obtido para a métrica $N1$ indica uma sobreposição significativa entre as classes, aspecto também evidenciado pelo índice de *silhouette*. A métrica de *separabilidade linear* apresentou um valor de 0,695, o que indica que, apesar da sobreposição, a classe apresenta um certo grau de separabilidade linear. Esse comportamento é compatível com a decomposição por PCA ilustrada na Figura 5.13, na qual é possível observar dois grandes aglomerados correspondentes às classes positiva e negativa, embora com a presença expressiva de amostras da classe oposta em cada agrupamento. Os valores estão listados na Tabela 5.7.

Tabela 5.7: Medidas de complexidade para a classe *hiper celularidade*.

Métrica	Valor
$N1$	0,442($\pm 0,015$)
Separabilidade Linear	0,694($\pm 0,018$)
Silhouette	0,365($\pm 0,027$)

Os resultados evidenciam que o processo de curadoria e o tratamento dos dados com ROS permitiram construir modelos com maior desempenho nos experimentos, superando aqueles treinados sem esses recursos. A lista dos modelos com as maiores taxas de F1-score é apresentada na Tabela 5.8.

Tabela 5.8: Taxas mais altas de F1-score alcançadas para *hiper celularidade*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas KCN, VGG-16 e EfficientNet-B0, sobre o conjunto ROS-Curated-HE.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(CNN)	ROS-Curated-HE	71,3% $\pm$ 2,1%	98,7% $\pm$ 2,0%	82,8% $\pm$ 2,0%
KCN _(CNN)	ROS-Curated-HE	67,5% $\pm$ 1,3%	94,5% $\pm$ 2,3%	78,8% $\pm$ 1,5%
KCN _(CNN)	Curated-HE	68,3% $\pm$ 1,1%	92,4% $\pm$ 1,0%	78,5% $\pm$ 1,0%
EfficientNet-B0 _(FSL)	Curated-HE	67,5% $\pm$ 1,5%	92,3% $\pm$ 1,6%	78,0% $\pm$ 1,1%
VGG-16 _(CNN)	ROS-Curated-HE	68,1% $\pm$ 0,9%	89,8% $\pm$ 1,4%	77,4% $\pm$ 0,8%
VGG-16 _(CNN)	Curated-HE	68,1% $\pm$ 1,3%	89,1% $\pm$ 1,6%	77,2% $\pm$ 1,3%
KCN _(CNN)	ROS-HE	65,2% $\pm$ 1,0%	93,2% $\pm$ 0,9%	76,8% $\pm$ 0,9%
Inception-v3 _(FSL)	Curated-HE	66,8% $\pm$ 1,2%	89,9% $\pm$ 5,0%	76,6% $\pm$ 2,2%
EfficientNet-B0 _(CNN)	ROS-Curated-HE	65,6% $\pm$ 1,3%	91,8% $\pm$ 1,7%	76,5% $\pm$ 0,7%
ResNet-18 _(FSL)	Curated-HE	66,1% $\pm$ 2,8%	91,0% $\pm$ 5,6%	76,5% $\pm$ 2,3%
KCN _(CNN)	HE	65,5% $\pm$ 0,8%	91,8% $\pm$ 1,3%	76,4% $\pm$ 0,6%

No que se refere à significância estatística dos resultados, o teste de postos com sinais de Wilcoxon não indicou evidência de diferença entre os desempenhos de modelos convencionais e aqueles baseados em redes prototípicas. Nos demais cenários comparativos, nos quais o teste não foi aplicado por não se tratar de modelos treinados sobre os mesmos conjuntos de dados, observou-se que o treinamento com dados refinados por curadoria resultou em desempenho superior ao obtido com dados não

tratados. Já na comparação entre modelos treinados com amostras tingidas de diferentes corantes e aqueles treinados com dados segmentados por corante, verificou-se que os modelos treinados com dados restritos a um único corante apresentaram desempenho superior.

A análise de predições incorretas aponta falsos positivos em todas as classes do conjunto, com maioria absoluta na classe *esclerose pura sem crescente*. A distribuição é exibida na Figura 5.14, enquanto as Figuras 5.15 e 5.16 exibem algumas amostras extraídos dos falsos positivos e falsos negativos.

Média das Predições Incorretas por Classe (Hiper celularidade)

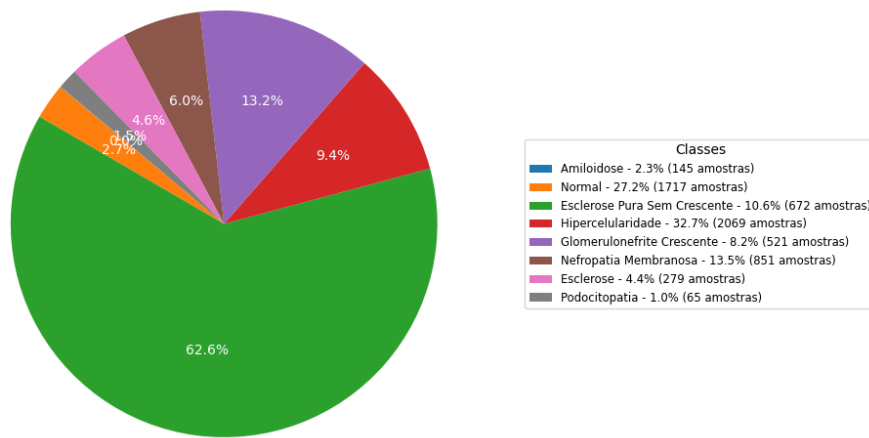


Figura 5.14: Distribuição das predições incorretas para *hiper celularidade*.

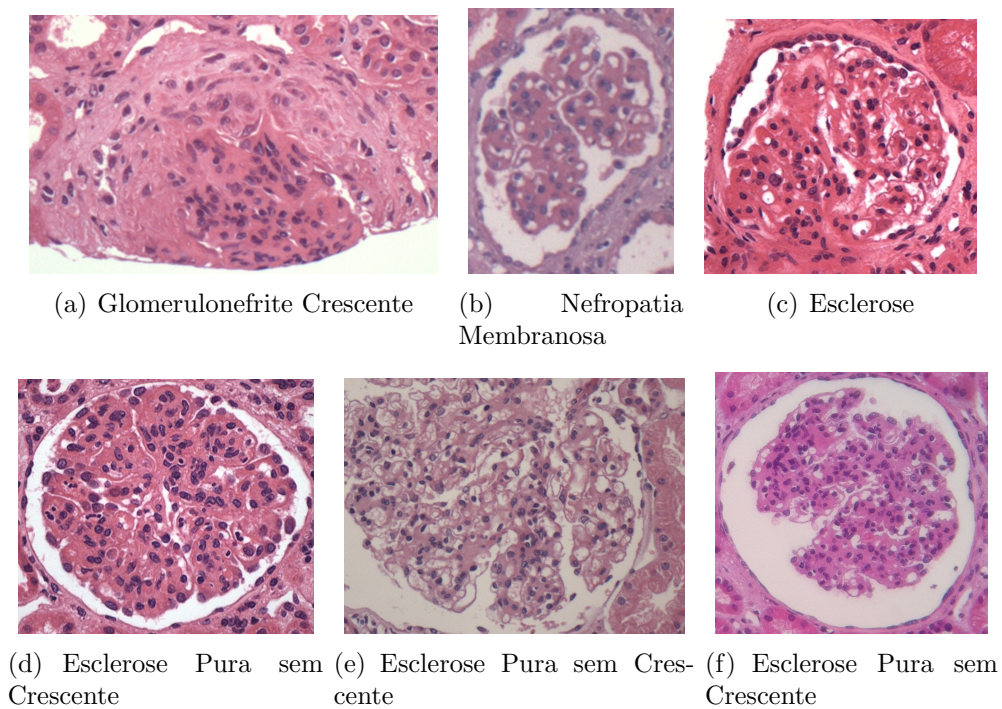


Figura 5.15: Amostras que produziram falsos positivos.

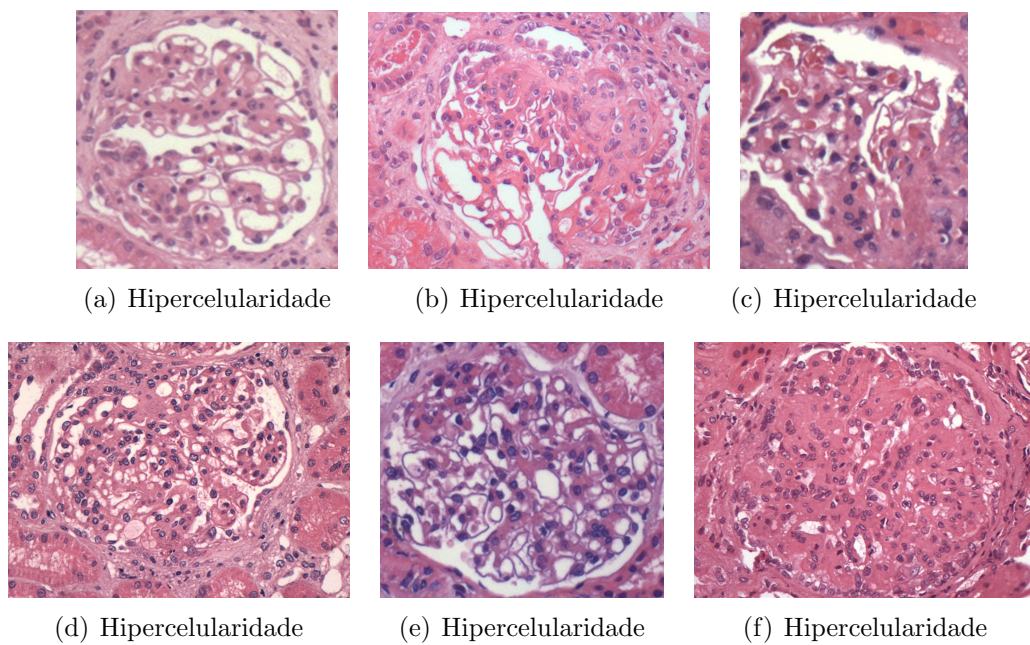


Figura 5.16: Amostras que produziram falsos negativos.

5.5 Resultados e Análise para a Classe Hiper celularidade Pura sem Crescente

A *hipercelularidade pura sem crescente* é a classe com menor representação em todo o conjunto estudado. Além disso, esta classe permaneceu quase totalmente inalterada após o processo de curadoria, com apenas uma amostra editada e ainda apresentando um aumento na taxa de desbalanceamento no conjunto refinado.

A análise de complexidade para *hipercelularidade pura sem crescente* apontou para um conjunto com alta complexidade. A Figura 5.17 trás a decomposição PCA para *hipercelularidade pura sem crescente* e é possível observar que essa classe encontra-se quase totalmente no e espaço de características da classe negativa, apresentando um desafio à sua correta separação.

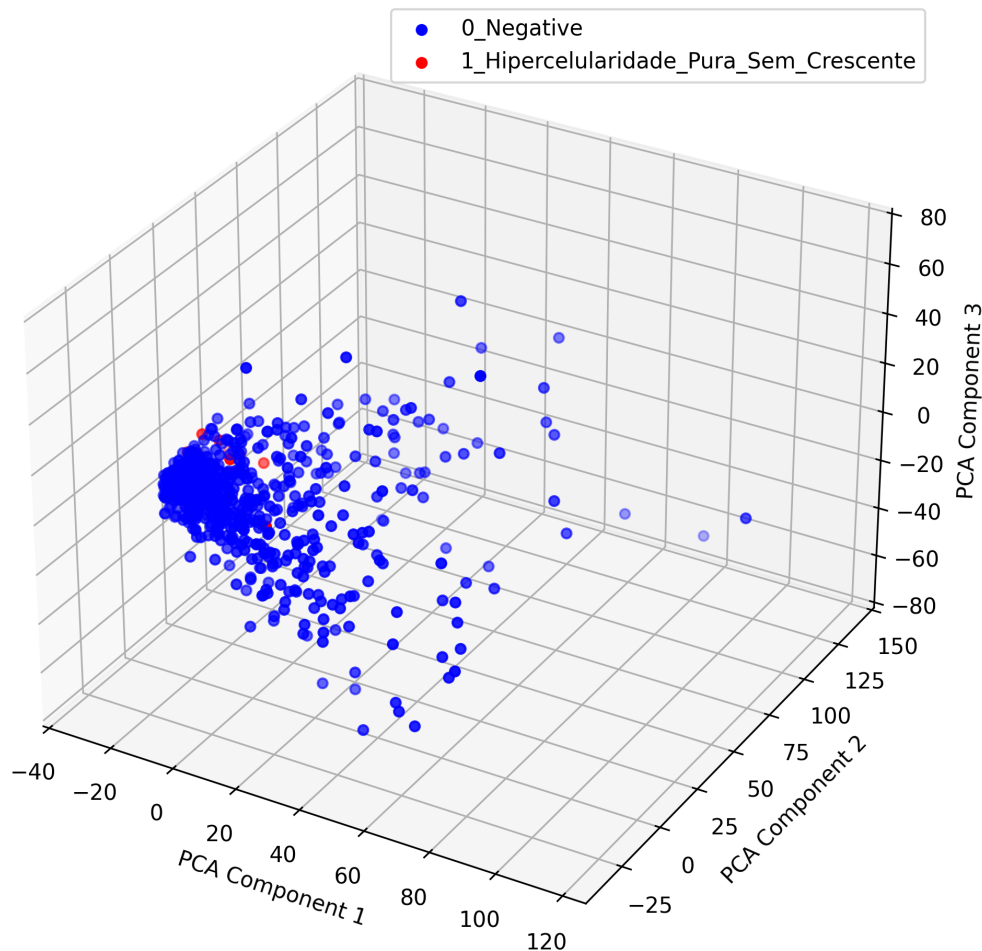


Figura 5.17: Análise de componentes principais para *hipercelularidade pura sem crescente*.

A análise ainda obteve o maior valor para $N1$ dentre as classes estudadas, indicando uma fronteira de decisão com sobreposição quase total entre as classes. A classe também não apresenta um espaço de características separável linearmente e o índice de *silhouette* apontou baixa compactação. Os resultados são exibidos na Tabela 5.9.

Tabela 5.9: Medidas de complexidade para a classe *hiper celularidade pura sem crescente*.

Métrica	Valor
$N1$	0,970($\pm 0,023$)
Separabilidade Linear	0,233($\pm 0,047$)
Silhouette	0,143($\pm 0,036$)

Os aspectos de complexidade mencionados podem ter contribuído para o baixo desempenho apresentado pelos classificadores, que permaneceu sempre abaixo de 30% de F1-score para os modelos individuais, com o modelo *ensemble-based* atingindo 34,42%($\pm 4,88\%$), como observado na Tabela 5.10.

Tabela 5.10: Taxas mais altas de F1-score alcançadas para *hiper celularidade pura sem crescente*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas VGG-16, Inception-v3 e ResNet-18, sobre o conjunto ROS-PAS.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(CNN)	ROS-PAS	22,0% $\pm$ 3,2%	90,3% $\pm$ 10,1%	34,4% $\pm$ 4,9%
VGG-16 _(CNN)	ROS-PAS	17,0% $\pm$ 0,7%	76,2% $\pm$ 7,9%	27,8% $\pm$ 1,4%
EfficientNet-B0 _(CNN)	PAS	17,4% $\pm$ 2,2%	68,9% $\pm$ 8,3%	27,7% $\pm$ 3,0%
EfficientNet-B0 _(FSL)	PAS	16,6% $\pm$ 2,4%	60,4% $\pm$ 16,1%	25,9% $\pm$ 4,3%
Inception-v3 _(CNN)	PAS	16,8% $\pm$ 2,9%	54,8% $\pm$ 7,6%	25,8% $\pm$ 4,1%
EfficientNet-B4 _(FSL)	PAS	15,6% $\pm$ 2,0%	75,0% $\pm$ 11,9%	25,7% $\pm$ 3,0%
ResNet-18 _(CNN)	PAS	16,8% $\pm$ 1,5%	54,0% $\pm$ 12,5%	25,7% $\pm$ 3,0%
Inception-v3 _(CNN)	ROS-PAS	14,7% $\pm$ 1,6%	79,9% $\pm$ 4,8%	24,9% $\pm$ 2,4%
VGG-16 _(CNN)	ROS	15,1% $\pm$ 0,8%	71,0% $\pm$ 9,5%	24,8% $\pm$ 1,0%
VGG-16 _(CNN)	ROS-Curated-PAS	14,9% $\pm$ 1,6%	66,5% $\pm$ 7,6%	24,4% $\pm$ 2,6%
EfficientNet-B0 _(FSL)	Curated-PAS	14,2% $\pm$ 2,2%	84,4% $\pm$ 11,4%	24,1% $\pm$ 2,7%

De forma semelhante ao ocorrido com a *esclerose pura sem crescente*, a degradação para a métrica F1-score foi causada pelo alto índice de falsos positivos, o que é evidenciado pelos notadamente baixos valores para *precision*. O gráfico exibido na Figura 5.18 descreve a distribuição das predições incorretas, com apenas 6,2% de falsos negativos. Ressalta-se o fato de que a maioria dos falsos positivos encontram-se na classe *hiper celularidade*, considerando a relação entre essas duas lesões, e ainda o alto índice de predições incorretas para *esclerose pura sem crescente*. As Figuras 5.19 e 5.20 exibem amostras extraídas dos erros de classificação.

Média das Predições Incorretas por Classe (Hiper celularidade Pura Sem Crescente)

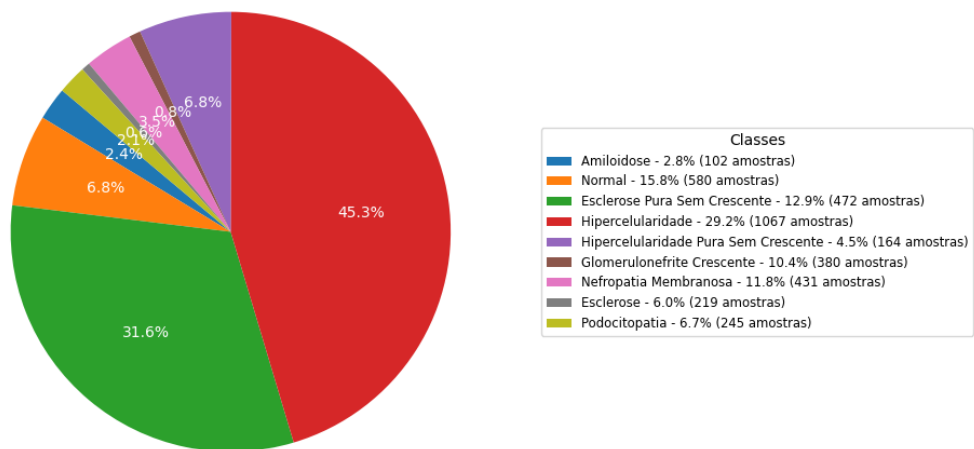


Figura 5.18: Distribuição das predições incorretas para *hiper celularidade pura sem crescente*.

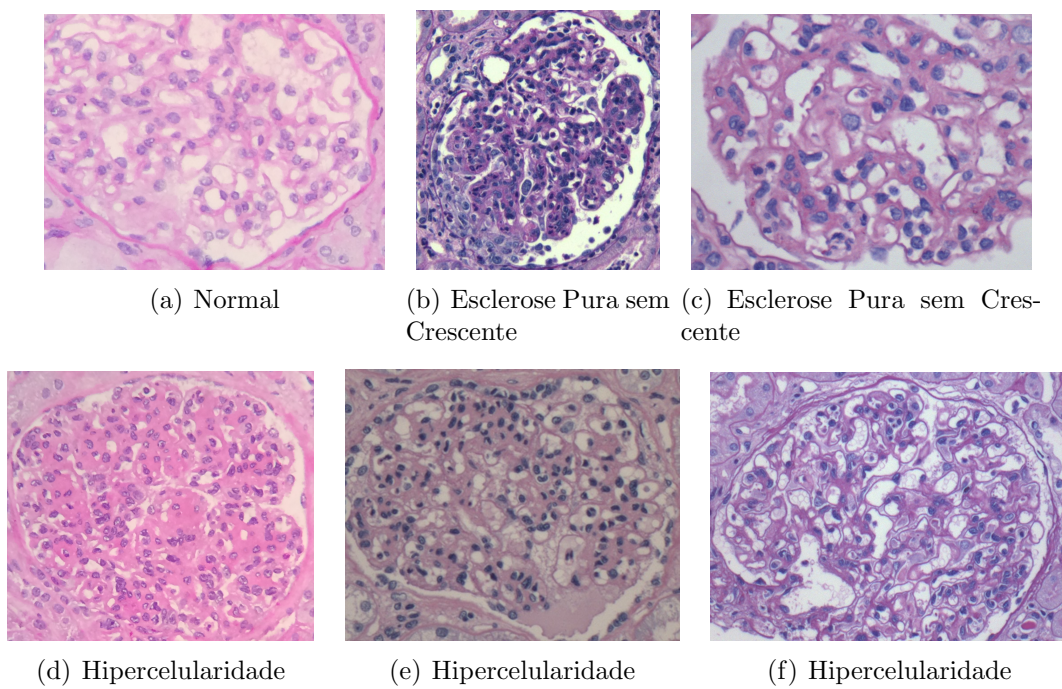


Figura 5.19: Amostras que produziram falsos positivos.

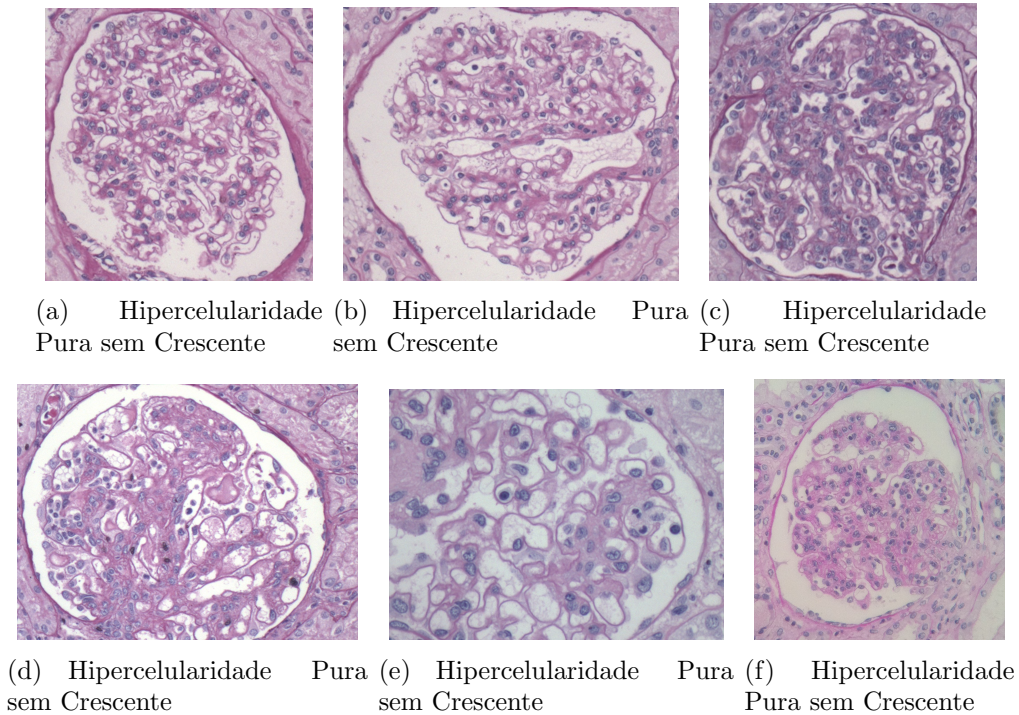


Figura 5.20: Amostras que produziram falsos negativos.

Em se tratando do comparativo entre técnicas comuns e FSL, não houve predominância absoluta de nenhum dos métodos, o que é corroborado com o que foi obtido no teste de significância estatística, com um resultado para o teste de Wilcoxon de 0,312. Os resultados mais altos para F1-score foram obtidos sobre os conjuntos restritos ao corante PAS. Vale salientar o fato de que a *hipercelularidade pura sem crescente* não possui amostras tingidas com HE, dentro do conjunto aqui utilizado. O ligeiro aumento do desempenho pode ter sido alcançado pela redução, ainda que baixa, da complexidade do conjunto, ao se impor uma segmentação por corante.

Os resultados para *hipercelularidade pura sem crescente* foram acentuadamente mais baixos do que aqueles obtidos para as outras classes. Tal observação sugere que esta classe seja analisada de forma separada para melhor entendimento das causas para o baixo desempenho.

5.6 Resultados e Análise para a Classe Glomerulonefrite Crescente

A *glomerulonefrite crescente* constitui uma classe com desbalanceamento moderado, com uma taxa aproximada de 1 : 10, e teve um acréscimo de 105 novas amostras após o processo de curadoria. A análise de complexidade do conjunto *Curated-glomerulonefrite* sugere uma classe com complexidade moderada. A decomposição

PCA exibida na Figura 5.21, indica um conjunto parcialmente separado, com algumas amostras muito próximas do espaço de características da classe negativa.

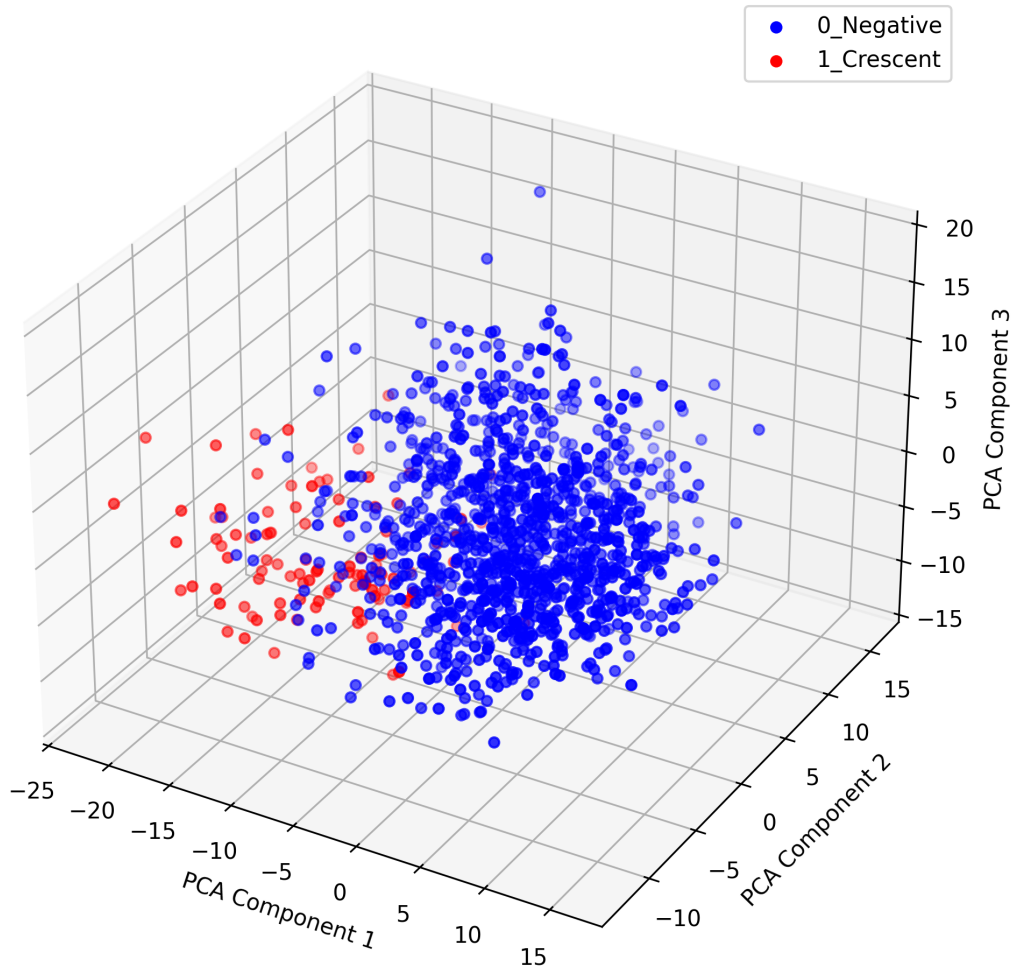


Figura 5.21: Análise de componentes principais *glomerulonefrite crescente*.

Para as outras medidas de complexidade do conjunto, observa-se certa *separabilidade linear*, com 70,6%. No entanto, o índice de *silhouette* sugere alta sobreposição e baixa compactação da classe de interesse. O valor obtido para $N1$ indica uma fronteira de decisão com sobreposição moderada. Os resultados são listados na Tabela 5.11.

Tabela 5.11: Medidas de complexidade para a classe *glomerulonefrite crescente*.

Métrica	Valor
N1	0,306($\pm 0,027$)
Separabilidade Linear	0,706($\pm 0,014$)
Silhouette	0,082($\pm 0,015$)

O desempenho obtido nos experimentos manteve-se consistentemente em torno de 80% para a métrica F1-score, sendo a maior taxa registrada sobre dados curados, sem aplicação de superamostragem, sem restrição por tipo de corante e utilizando um modelo *ensemble-based*, conforme apresentado na Tabela 5.12. Vale destacar que, embora o modelo *ensemble-based* tenha superado em 7,38% o maior desempenho obtido por um modelo individual, observa-se um desvio padrão relativamente elevado, especialmente quando comparado aos desvios observados nos demais modelos, o que reduz a robustez do resultado.

Tabela 5.12: Taxas mais altas de F1-score alcançadas para *glomerulonefrite crescente*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas VGG-16, Inception-v3 e ResNet-18, sobre o conjunto *Curated-glomerulonefrite*.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(CNN)	Curated	88,7% $\pm$ 5,8%	94,3% $\pm$ 4,6%	91,4% $\pm$ 4,9%
Inception-v3 _(CNN)	Curated	82,1% $\pm$ 3,3%	86,3% $\pm$ 3,1%	84,0% $\pm$ 0,7%
Inception-v3 _(CNN)	Curated-HE	80,1% $\pm$ 2,9%	87,0% $\pm$ 0,9%	83,7% $\pm$ 1,5%
ResNet-50 _(FSL)	Curated-PAS	84,3% $\pm$ 4,3%	82,9% $\pm$ 2,6%	83,6% $\pm$ 3,4%
Inception-v3 _(CNN)	ROS-Curated	80,7% $\pm$ 2,3%	86,6% $\pm$ 1,1%	83,5% $\pm$ 1,3%
Inception-v3 _(FSL)	Curated-HE	82,3% $\pm$ 4,5%	84,6% $\pm$ 3,8%	83,3% $\pm$ 1,5%
ResNet-18 _(CNN)	Curated	79,5% $\pm$ 1,7%	87,3% $\pm$ 1,3%	83,2% $\pm$ 1,2%
Inception-v3 _(CNN)	ROS-Curated-HE	81,9% $\pm$ 5,1%	84,3% $\pm$ 3,6%	83,0% $\pm$ 3,7%
VGG-16 _(CNN)	ROS-Curated	78,1% $\pm$ 1,8%	88,0% $\pm$ 1,5%	82,7% $\pm$ 0,7%
ResNet-18 _(CNN)	Curated-HE	78,5% $\pm$ 4,6%	87,3% $\pm$ 2,1%	82,6% $\pm$ 3,4%
Inception-v3 _(CNN)	Curated-PAS	80,5% $\pm$ 4,7%	84,5% $\pm$ 4,4%	82,3 $\pm$ 4,4%

Embora os resultados com FSL tenham ficado próximos ao obtido com métodos comuns, especialmente no tocante aos experimentos realizados em conjuntos restritos a um corante específico, teste de significância estatística apontou a superioridade dos modelos baseados em CNN, com um valor de 0,062. A técnica alcançou o desempenho de 83,58%($\pm 3,35\%$) no conjunto *Curated-PAS-glomerulonefrite* e 83,31%($\pm 1,53\%$) no conjunto *Curated-HE-glomerulonefrite*, que apresentam respectivamente, taxas de desbalanceamento de $\approx 1 : 9$ e $\approx 1 : 11$. A menor taxa de desbalanceamento no conjunto restrito ao PAS, pode justificar a pequena superioridade no desempenho em relação ao obtido no conjunto restrito ao HE, porém, salienta-se o fato de que o desvio padrão para esses experimentos também foi maior, o que indica um aprendizado mais estável dos modelos treinados com amostras tingidas com HE.

A análise de predições incorretas apontou que *glomerulonefrite crescente* apresentou falsos positivos distribuídos em todas as classes do conjunto, com sua maioria presente na classe *hipercelularidade*. A distribuição completa é exibida na Figura 5.22, enquanto as Figuras 5.23 e 5.24 exibem algumas amostras apontadas nos erros de classificação.

Média das Predições Incorretas por Classe (Glomerulonefrite Crescente)

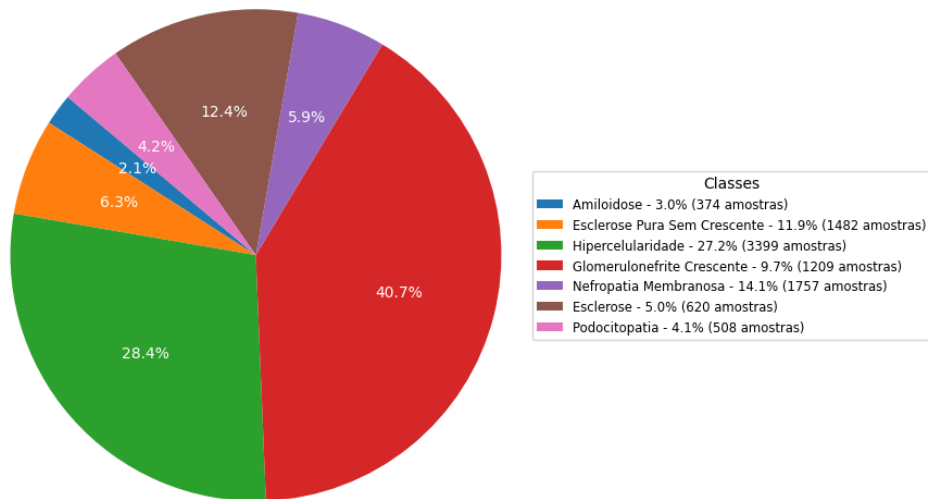
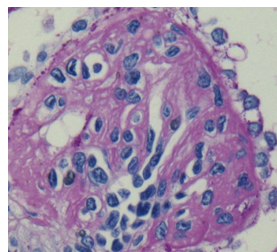
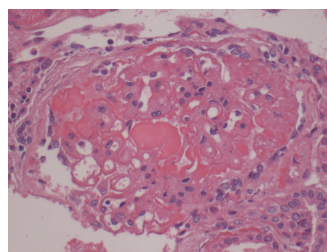


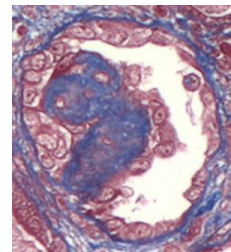
Figura 5.22: Distribuição das predições incorretas para *glomerulonefrite crescente*.



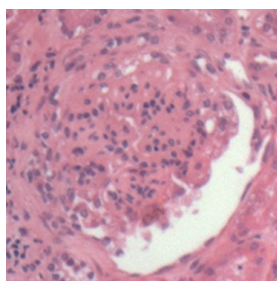
(a) Esclerose Pura sem Crescente



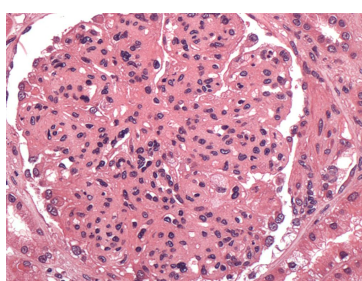
(b) Esclerose



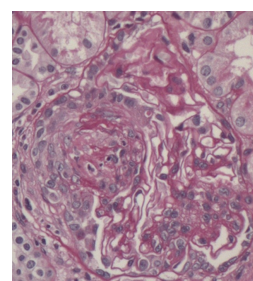
(c) Podocitopatia



(d) Hipercelularidade



(e) Hipercelularidade



(f) Hipercelularidade

Figura 5.23: Amostras que produziram falsos positivos.

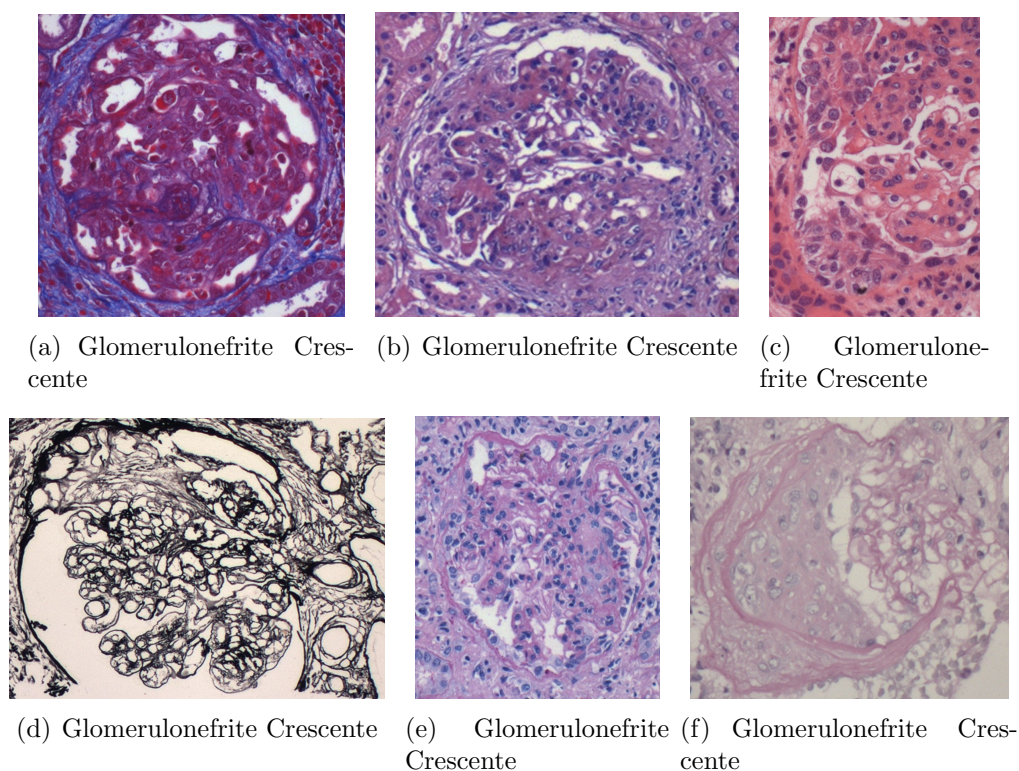


Figura 5.24: Amostras que produziram falsos negativos.

5.7 Resultados e Análise para a Classe Nefropatia Membranosa

A *nefropatia membranosa* é uma classe taxa de desbalanceamento moderada, variando entre $\approx 1 : 6$ a $\approx 1 : 8$ entre os conjuntos utilizados neste estudo. A análise de complexidade dos dados apontou para um conjunto com complexidade moderada. A decomposição PCA exibida na Figura 5.25 revela uma classe positiva parcialmente separada, com muitas amostras próximas ou mesmo contidas no espaço de características da classe negativa.

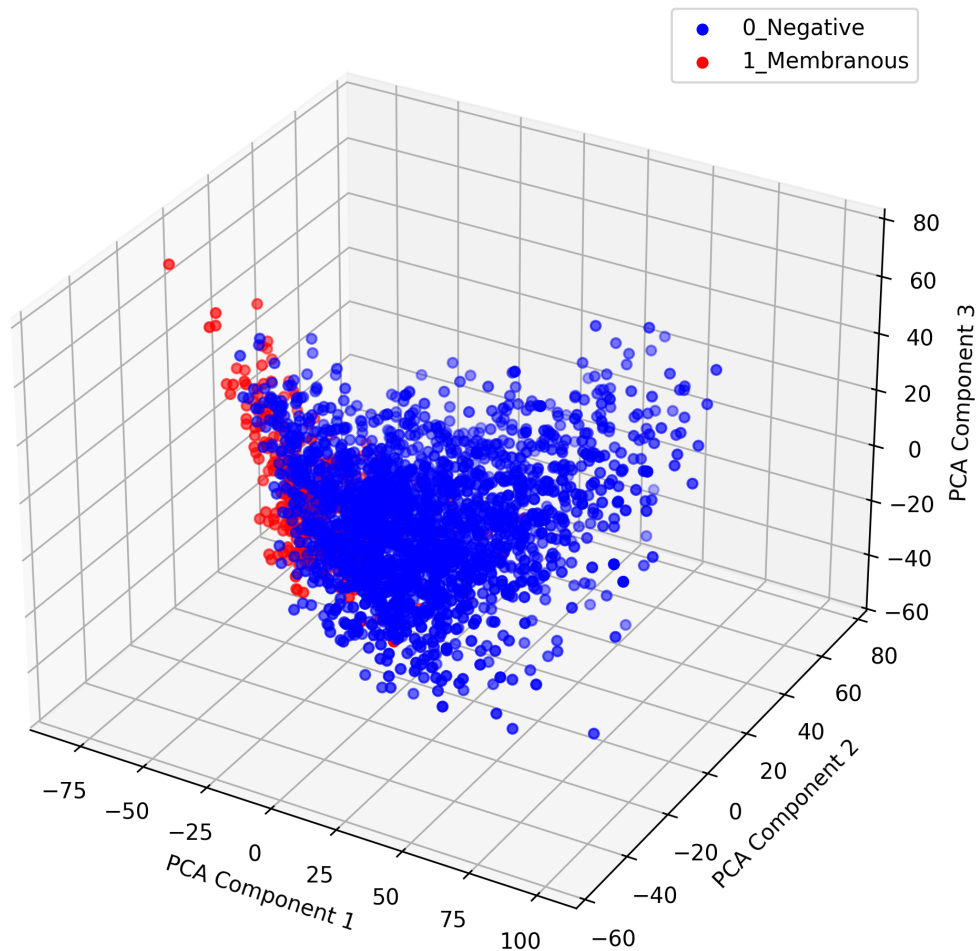


Figura 5.25: Análise de componentes principais para *nefropatia membranosa*.

A métrica de linearidade sugere uma classe com certa *separabilidade*. Já os índices de *silhouette* e *N1* apontam para uma classe com baixa compactação e sobreposição moderada. Os resultados da análise são listados na Tabela 5.13.

Tabela 5.13: Medidas de complexidade para a classe *nefropatia membranosa*.

Métrica	Valor
N1	0,341($\pm 0,020$)
Separabilidade Linear	0,637($\pm 0,048$)
Silhouette	0,117($\pm 0,020$)

O desempenho dos modelos, com poucas exceções, permaneceu acima dos 86% de F1-score, chegando a ultrapassar os 90%, com o modelo *ensemble-based* construído com redes prototípicas, sobre dados curados, restritos ao corante HE, como mostra

a Tabela 5.14. Salienta-se neste resultado o desvio padrão relativamente alto obtido pelo modelo baseado em conjunto (5,17%).

Tabela 5.14: Taxas mais altas de F1-score alcançadas para *nefropatia membranosa*. O modelo *ensemble-based* foi construído pelos modelos produzidos pelas arquiteturas Inception-v3, ResNet-50 e ResNet-18, sobre o conjunto *Curated-HE-Membranosa*.

Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(FLS)	Curated-HE	84,9% $\pm$ 8,0%	99,3% $\pm$ 1,6%	91,4% $\pm$ 5,2%
Inception-v3 _(FLS)	Curated-HE	86,0% $\pm$ 7,5%	90,0% $\pm$ 2,1%	87,7% $\pm$ 3,2%
ResNet-50 _(FLS)	Curated-HE	83,0% $\pm$ 4,6%	92,4% $\pm$ 1,3%	87,3% $\pm$ 2,5%
ResNet-50 _(FLS)	Curated-PAS	89,3% $\pm$ 5,3%	85,2% $\pm$ 3,9%	87,1% $\pm$ 3,6%
ResNet-18 _(FLS)	Curated-PAS	84,7% $\pm$ 6,1%	88,9% $\pm$ 1,9%	86,7% $\pm$ 3,3%
VGG-16 _(CNN)	ROS-Curated	84,7% $\pm$ 2,1%	88,6% $\pm$ 2,0%	86,6% $\pm$ 1,8%
Inception-v3 _(CNN)	Curated-HE	82,9% $\pm$ 5,5%	90,8% $\pm$ 2,5%	86,6% $\pm$ 2,5%
Inception-v3 _(FLS)	Curated-PAS	86,9% $\pm$ 6,3%	86,3% $\pm$ 2,2%	86,5% $\pm$ 3,0%
Inception-v3 _(CNN)	Curated	83,3% $\pm$ 2,9%	89,7% $\pm$ 1,2%	86,3% $\pm$ 1,1%
ResNet-18 _(CNN)	Curated-HE	83,2% $\pm$ 2,1%	89,4% $\pm$ 1,4%	86,2% $\pm$ 1,6%
Inception-v3 _(CNN)	ROS-Curated-HE	85,4% $\pm$ 3,7%	87,1% $\pm$ 3,4%	86,1% $\pm$ 1,0%

Utilizando-se os métodos comuns e modelos baseados em CNN, o resultado mais alto, para um modelo individual, foi de 86,60($\pm$ 1,82%), utilizando dados curados e com superamostragem ROS. Entretanto, salienta-se que não ficou evidenciado que este resultado é estatisticamente superior ao obtido com FSL, já que o teste de Wilcoxon apresentou um resultado de 0,312. Um fato relevante a ser mencionado é que o processo de curadoria rendeu a essa classe 218 novas amostras, baixando a taxa de desbalanceamento geral do conjunto de aproximadamente 1 : 7 para cerca de 1 : 6, possivelmente contribuindo para o aumento do desempenho observado nos experimentos.

A análise de predições incorretas aponta que a maioria dos falsos positivos ocorreu na classe *normal*, seguida por *hipercelularidade*. Entretanto, observa-se também um numero significativo de falsos positivos para as outras classes, com exceção de *hipercelularidade pura sem crescente*, e uma taxa de 48,9% de falsos negativos. A distribuição é exibida na Figura 5.26. As Figuras 5.27 e 5.28 exibem amostras extraídas dos erros classificação.

Média das Predições Incorretas por Classe (Nefropatia Membranosa)

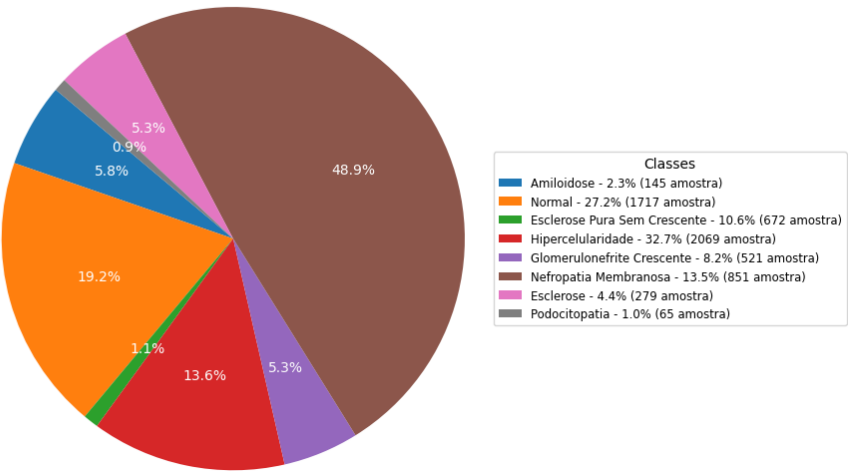


Figura 5.26: Distribuição das predições incorretas para *nefropatia membranosa*.

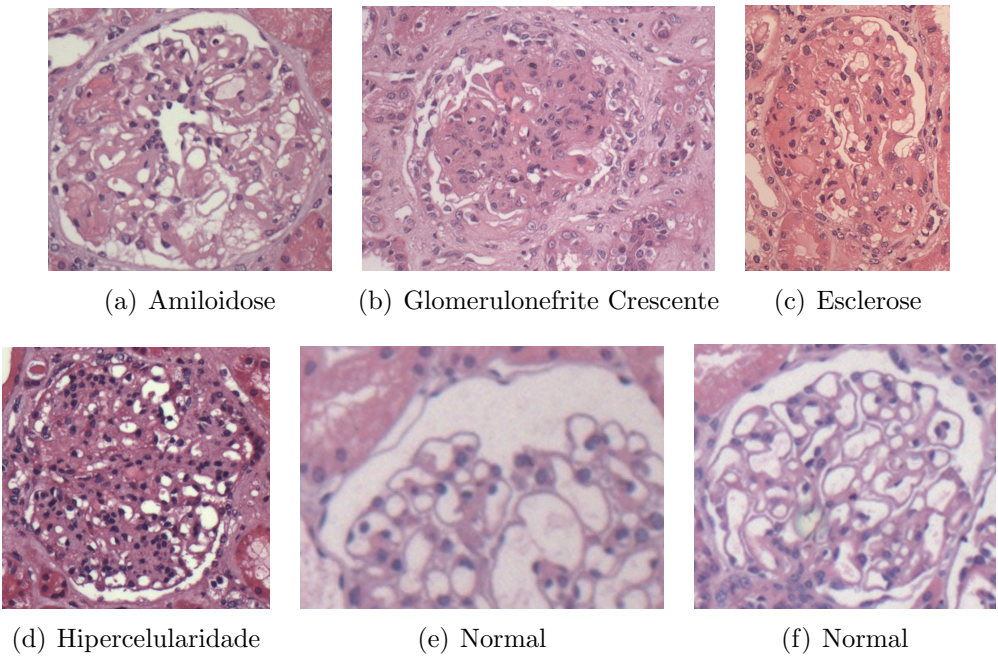


Figura 5.27: Amostras que produziram falsos positivos.

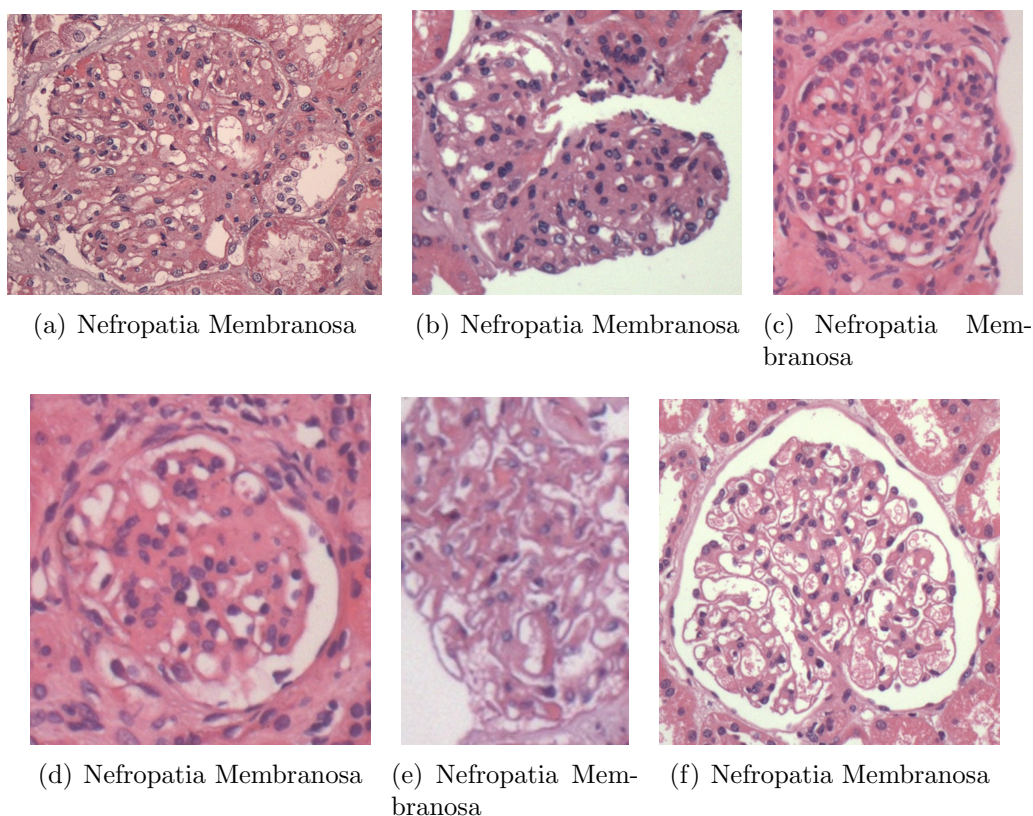


Figura 5.28: Amostras que produziram falsos negativos.

5.8 Resultados e Análise para a Classe Esclerose

No conjunto estudado, a classe *esclerose* apresenta um desbalanceamento acentuado, com uma razão aproximada de 1:22 nos subconjuntos restritos ao corante HE. O processo de curadoria resultou na adição de apenas três novas amostras, o que não foi suficiente para atenuar o elevado grau de desbalanceamento.

Com respeito à análise de complexidade do conjunto, a decomposição PCA, descrita na Figura 5.29, indica uma classe positiva com amostras distribuídas no espaço vetorial, em regiões muito próximas ou contidas no espaço da classe negativa, o que fornece uma noção do desafio imposto pela classe em sua separação por parte dos modelos.

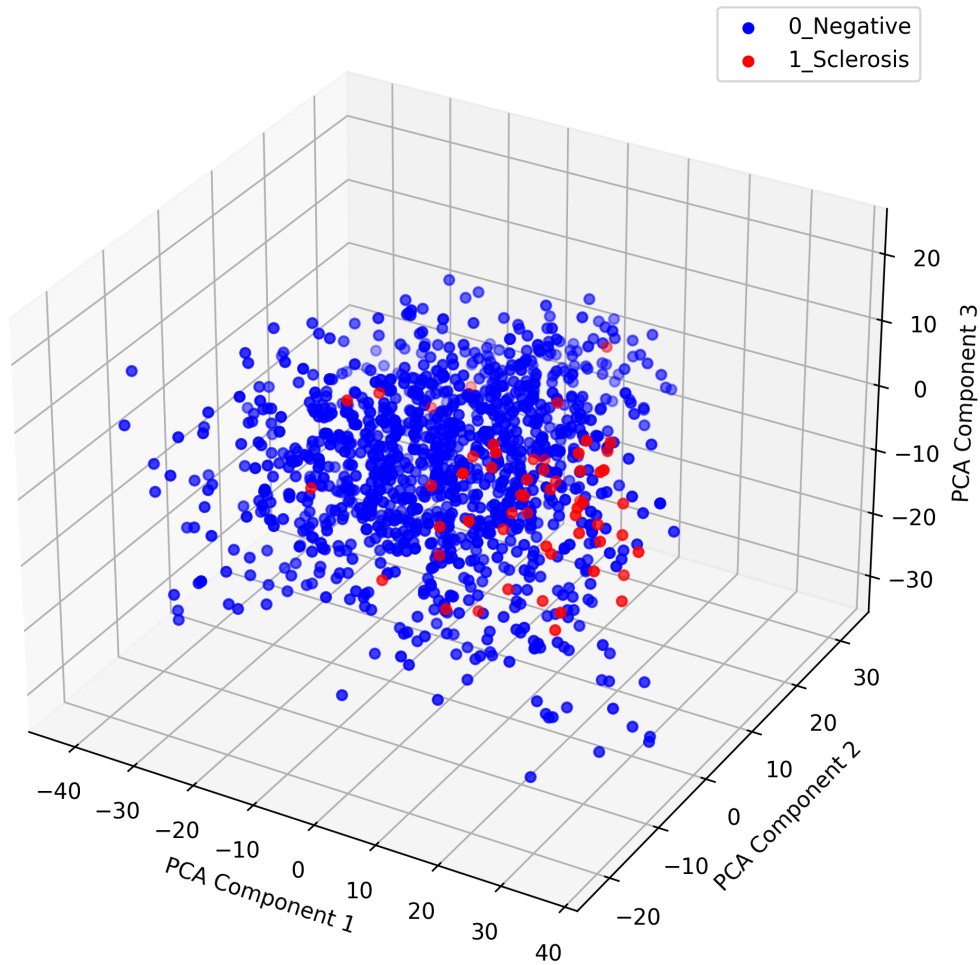


Figura 5.29: Análise de componentes principais para *esclerose*.

As demais métricas de complexidade apresentaram resultados que sugerem um conjunto com alta sobreposição e compactação moderada, o que pode ser inferido pelos valores observados para $N1$ e o índice de *silhouette*. Foi obtido também um baixo índice de *separabilidade linear*. Os resultados são listados na Tabela 5.15.

Tabela 5.15: Medidas de complexidade para a classe *esclerose*.

Métrica	Valor
$N1$	$0,582(\pm 0,060)$
Separabilidade Linear	$0,446(\pm 0,023)$
Silhouette	$0,453(\pm 0,044)$

O desempenho mais alto para um modelo individual foi de $73,91\%(\pm 7,93\%)$ de F1-

score. O modelo *ensemble-based* obteve 85,27%, entretanto, acompanhado de um desvio padrão elevado (14,10%), o que compromete a consistência e a confiabilidade do desempenho obtido. A lista dos maiores resultados é exibida na Tabela 5.16.

Tabela 5.16: Taxas mais altas de F1-score alcançadas para *esclerose*. O modelo *ensemble-based* foi construído com modelos produzidos pelas arquiteturas ResNet-50, ResNet-18 e Inception-v3, sobre o conjunto *PAS-esclerose*.

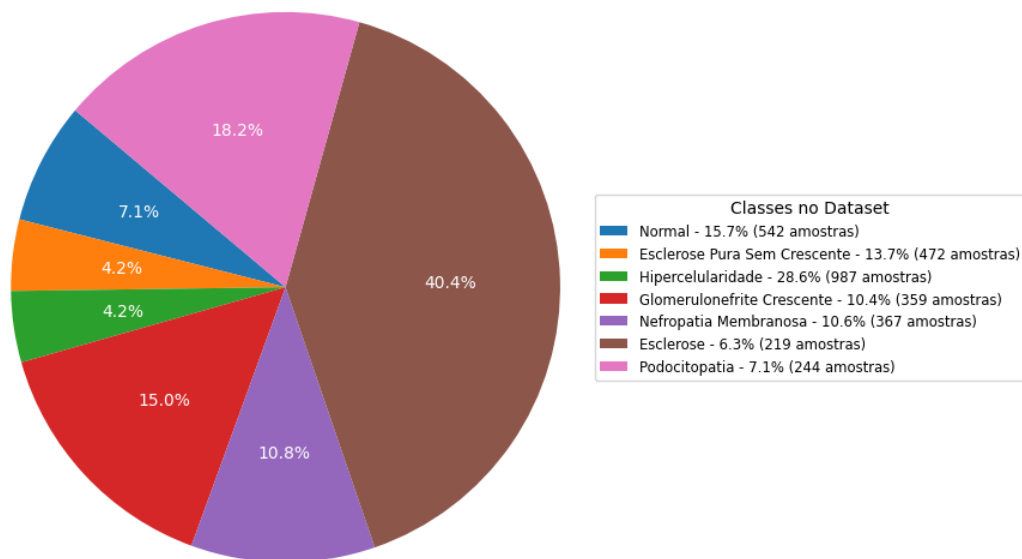
Classif.	Conjunto	Precisão	Recall	F1-Score
Ensemble _(FSL)	PAS	78,9% ± 15,8%	93,2% ± 11,6%	85,3% ± 14,1%
ResNet-50 _(FSL)	PAS	68,9% ± 11,8%	80,8% ± 6,6%	73,9% ± 7,9%
Inception-v3 _(FSL)	HE	72,2% ± 9,3%	75,0% ± 8,8%	73,3% ± 7,4%
ResNet-50 _(FSL)	HE	66,5% ± 6,1%	78,6% ± 4,9%	72,0% ± 5,5%
Inception-v3 _(FSL)	Curated-HE	71,2% ± 4,7%	73,1% ± 5,4%	71,9% ± 2,5%
Inception-v3 _(FSL)	PAS	69,5% ± 5,8%	74,4% ± 9,3%	71,3% ± 3,5%
ResNet-50 _(FSL)	Curated-HE	65,6% ± 2,6%	77,4% ± 5,0%	71,0% ± 3,5%
ResNet-18 _(FSL)	Curated-HE	68,0% ± 11,5%	74,2% ± 4,8%	70,4% ± 7,3%
ResNet-18 _(CNN)	Curated-HE	64,5% ± 4,6%	77,4% ± 4,7%	70,3% ± 4,1%
Inception-v3 _(CNN)	HE	61,7% ± 8,8%	81,9% ± 9,3%	70,2% ± 8,0%
ResNet-18 _(FSL)	HE	62,2% ± 3,9%	81,2% ± 11,3%	70,0% ± 4,9%

Como exibido na Tabela 5.16, o resultado mais alto foi alcançado sobre o conjunto *PAS-esclerose*, utilizando-se FSL, com uma taxa de F1-score de 73,91%(±7,93). Embora tenha sido observada nos experimentos uma superioridade das redes prototípicas sobre os métodos comuns, o resultado do teste de Wilcoxon não comprovou essa superioridade do ponto de vista estatístico com um resultado de 0,187.

Para as comparações feitas envolvendo conjuntos distintos, se observou vantagem no uso de conjuntos restritos por corante em comparação ao desempenho sobre dados sem restrição. Entretanto, é importante atentar-se para o elevado desvio padrão observado nos resultados, na maioria dos casos, ultrapassando os 4%, indicando uma inconsistência dos modelos, frente à variabilidade dos dados.

A análise de predições incorretas, exibida na Figura 5.30, indica a presença de falsos positivos em todas as classes do conjunto, com a maioria destes na classe *normal*, seguida da classe *hipercelularidade*. Essa distribuição condiz com o exibido na decomposição PCA, uma vez que se vê amostras da classe positiva distribuídas por diversas regiões do espaço vetorial.

Média das Predições Incorretas por Classe (Esclerose)

Figura 5.30: Distribuição das predições incorretas para *esclerose*.

Ao se analisar também as amostras envolvidas nos erros de predição, observou-se que uma parte significativa delas apresentava algum tipo de distorção previamente tratada no processo de curadoria, conforme ilustrado nas Figuras 5.31 e 5.32. Destaca-se, ainda, que embora o maior desempenho tenha sido alcançado com dados não curados, o melhor resultado obtido com um conjunto refinado apresentou o menor desvio padrão, sugerindo um desempenho mais consistente quando se utilizam dados curados.

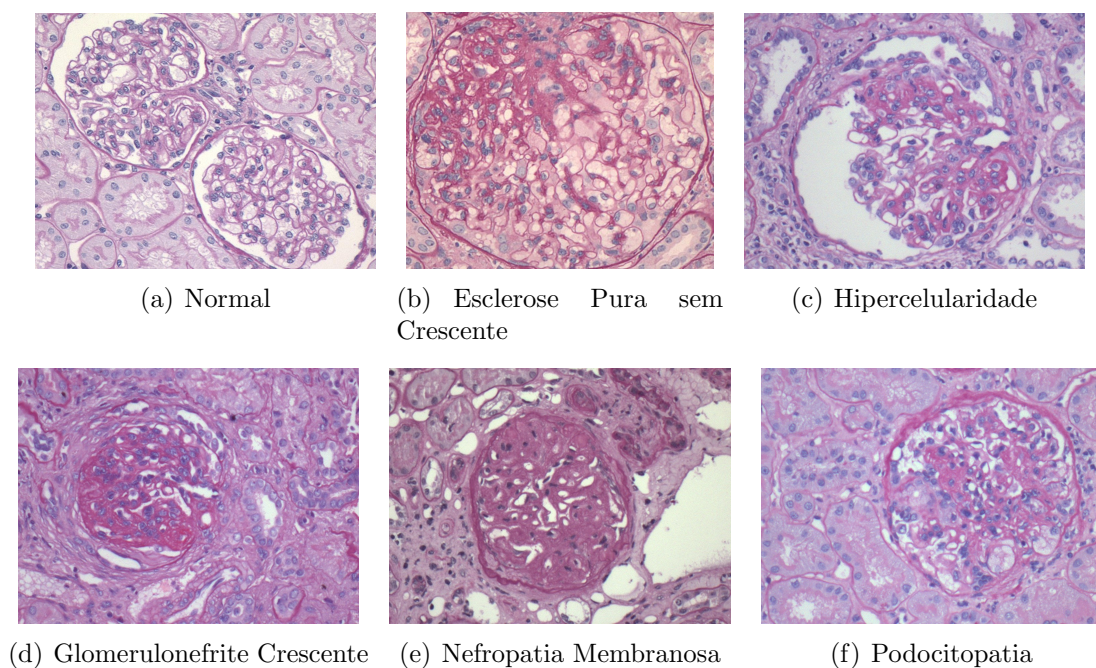


Figura 5.31: Amostras que produziram falsos positivos.

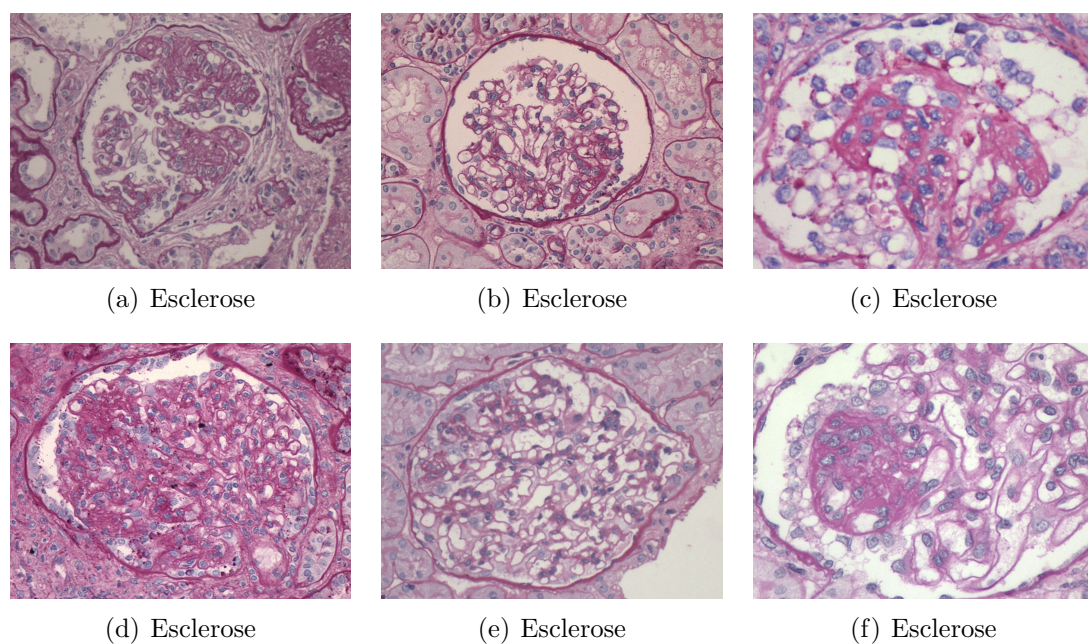


Figura 5.32: Amostras que produziram falsos negativos.

5.9 Resultados e Análise para a Classe Podocitopatía

A classe *podocitopatía* foi a classe que apresentou a maior taxa de desbalanceamento neste estudo, quando utilizados conjuntos curados restritos ao corante HE, chegando a $\approx 1 : 96$. Nesse cenário, a classe apresentou somente 65 amostras, o que apresentou um desafio para o treinamento dos modelos, especialmente aqueles que utilizaram métodos comuns. Esta classe também não recebeu o acréscimo de novas imagens no processo de curadoria, tendo apenas duas amostras editadas para a correção de distorções de ocupação.

Quanto à complexidade do conjunto, a análise de componentes principais, exibida na Figura 5.33, mostra uma classe de difícil separação, com elementos intrinsecamente mesclados entre amostras da classe negativa.

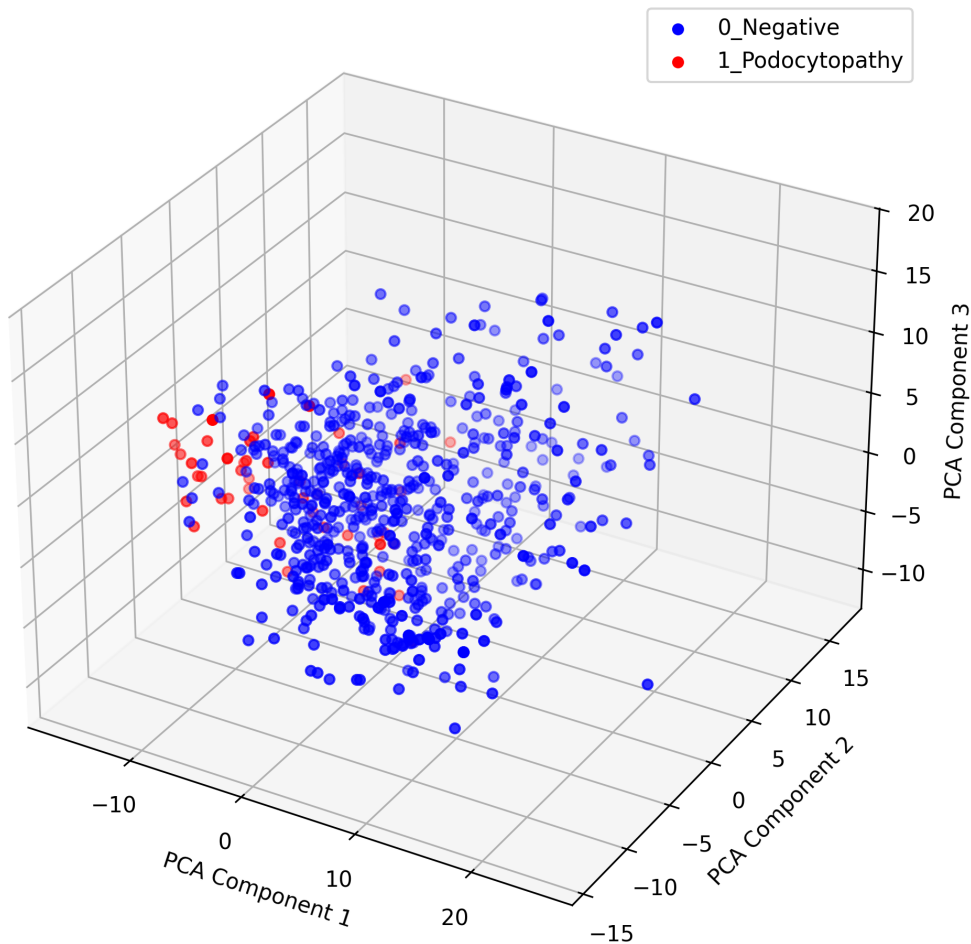


Figura 5.33: Análise de componentes principais para *podocitopatía*.

Para as demais métricas, a análise apontou um baixo índice de *separabilidade linear*.

Também ficou evidenciado que o conjunto de dados para *podocitopatia* apresenta alto índice de sobreposição e baixa compactação, o que é evidenciado pelos valores obtidos para *N1* e *silhouette*. Os resultados da análise são mostrados na Tabela 5.17.

Tabela 5.17: Medidas de complexidade para a classe *podocitopatia*. O modelo *ensemble-based* foi construído com os modelos produzidos pelas arquiteturas ResNet-50, ResNet-18 e Inception-v3, sobre o conjunto *PAS-podocitopatia*.

Métrica	Valor
N1	0,433($\pm 0,056$)
Separabilidade Linear	0,455($\pm 0,043$)
Silhouette	0,160($\pm 0,014$)

Os resultados para essa classe são apresentados na Tabela 5.18, onde se observa que o modelo *ensemble-based* obteve desempenho 0,17% inferior ao alcançado pelo modelo baseado em redes prototípicas sobre o conjunto *PAS-podocitopatia*, apresentando ainda um alto desvio padrão.

Tabela 5.18: Taxas mais altas de F1-score alcançadas para *podocitopatia*.

Classif.	Conjunto	Precisão	Recall	F1-Score
ResNet-18 _(FSL)	PAS	72,1% $\pm$ 4,0%	79,9% $\pm$ 8,1%	75,8% $\pm$ 5,8%
Ensemble _(FSL)	PAS	63,9% $\pm$ 21,3%	95,9% $\pm$ 9,1%	75,6% $\pm$ 17,7%
ResNet-18 _(CNN)	Baseline	70,1% $\pm$ 2,8%	80,4% $\pm$ 6,6%	74,9% $\pm$ 4,3%
ResNet-50 _(CNN)	Baseline	68,8% $\pm$ 1,8%	78,6% $\pm$ 4,3%	73,4% $\pm$ 2,3%
ResNet-18 _(CNN)	PAS	70,8% $\pm$ 5,7%	76,2% $\pm$ 7,7%	73,3% $\pm$ 5,8%
Inception-v3 _(CNN)	Baseline	69,0% $\pm$ 3,2%	78,4% $\pm$ 4,5%	73,3% $\pm$ 0,8%
Inception-v3 _(CNN)	PAS	66,2% $\pm$ 5,8%	75,8% $\pm$ 7,1%	70,6% $\pm$ 5,6%
Inception-v3 _(FSL)	PAS	67,7% $\pm$ 7,3%	71,7% $\pm$ 8,5%	69,3% $\pm$ 5,3%
ResNet-50 _(FSL)	PAS	63,6% $\pm$ 12,6%	77,0% $\pm$ 8,5%	69,0% $\pm$ 8,9%
Inception-v3 _(CNN)	Curated	62,5% $\pm$ 4,6%	76,5% $\pm$ 2,1%	68,7% $\pm$ 3,3%
ResNet-50 _(CNN)	Curated	63,8% $\pm$ 3,5%	73,7% $\pm$ 2,4%	68,3% $\pm$ 2,3%

O maior desempenho foi alcançado com o uso de redes prototípicas, aplicadas sobre o conjunto restrito ao corante PAS, com 75,76%($\pm 5,76\%$), porém, de modo geral, houve predominância dos modelos construídos com CNNs convencionais, com o segundo maior desempenho tendo sido obtido por um desses modelos, sobre o conjunto *Baseline-podocitopatia* (74,85%($\pm 4,30\%$)). Ressalta-se ainda que o menor desvio padrão também foi alcançado por modelo baseado em CNN, o que contrasta com os desvios-padrão mais elevados, obtidos com FSL. O teste de Wilcoxon apresentou um resultado de 0,312, indicando que não existe diferença estatisticamente significativa entre esses modelos.

A análise das predições incorretas revela a presença de falsos positivos distribuídos entre todas as classes, com predominância na classe *esclerose pura sem crescente*, seguida pela classe *esclerose*. Esse padrão está alinhado com o observado na análise

de componentes principais, na qual as amostras da classe positiva apresentam-se relativamente próximas ao espaço de características da classe negativa, indicando uma sobreposição entre os grupos que pode dificultar a discriminação pelos modelos. A distribuição é exibida no gráfico da Figura 5.34, enquanto as Figuras 5.35 e 5.36 exibem algumas amostras presentes nos erros de classificação.

Média das Predições Incorretas por Classe (Podocitopatia)

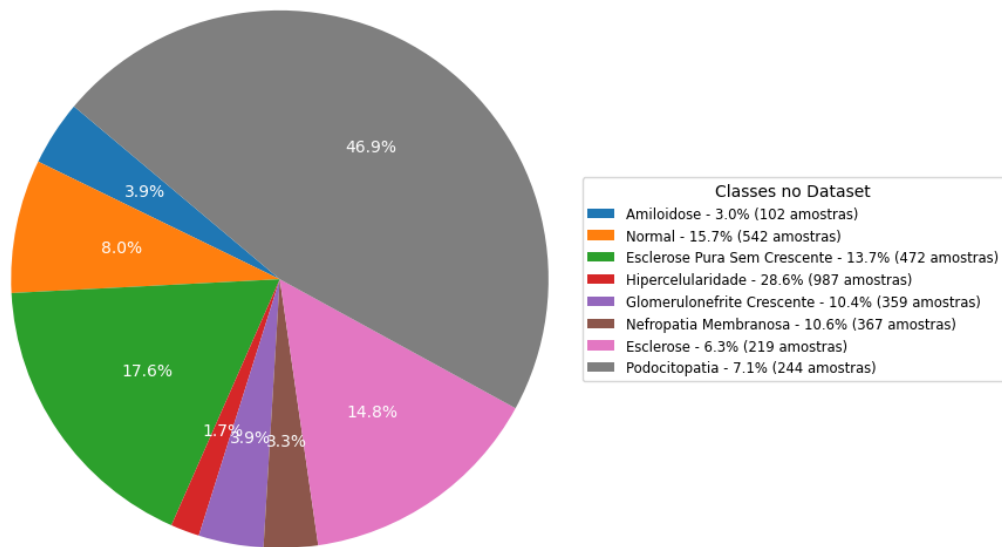


Figura 5.34: Distribuição das predições incorretas para *podocitopatia*.

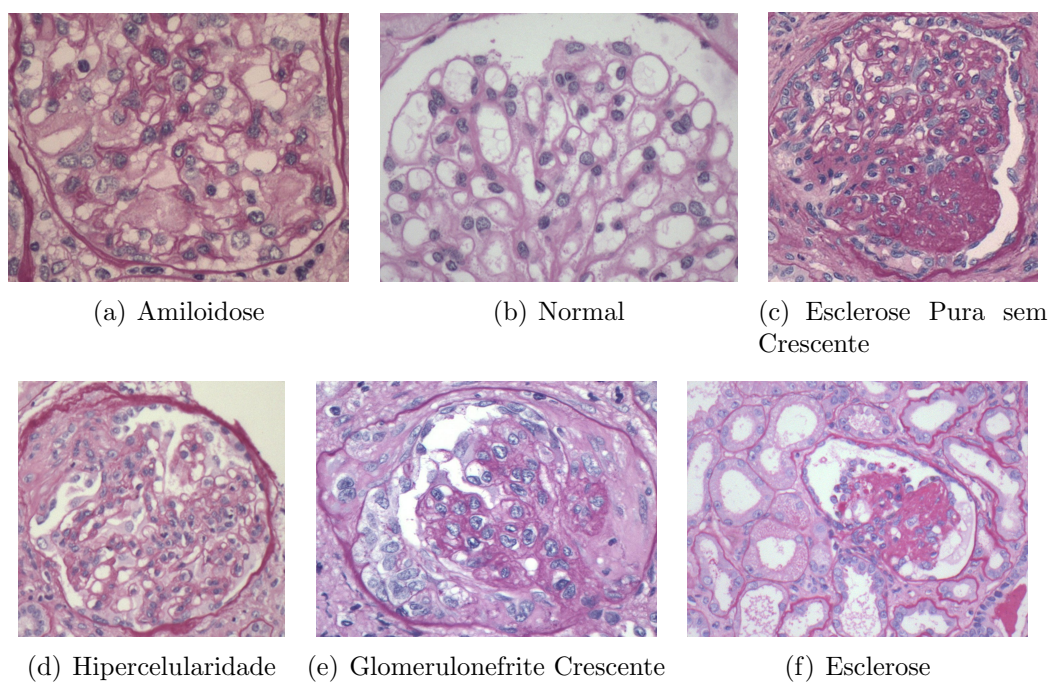


Figura 5.35: Amostras que produziram falsos positivos.

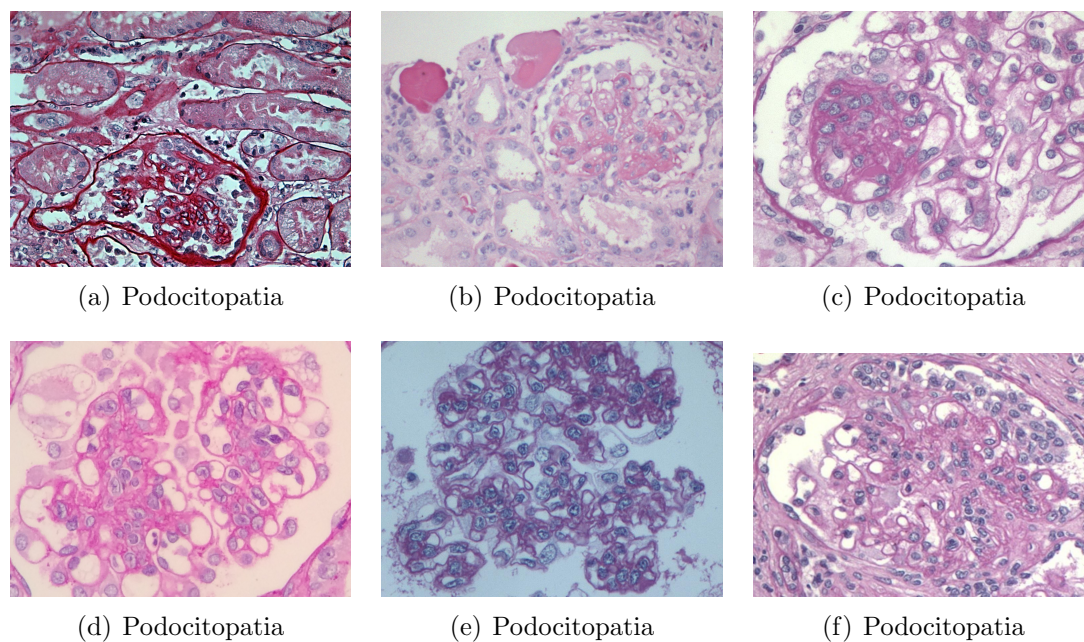


Figura 5.36: Amostras que produziram falsos negativos.

5.10 Análise dos Resultados

Observando-se os resultados obtidos em todos os experimentos, é possível inferir que a complexidade dos dados pode ter sido um fator mais crítico do que a própria taxa de desbalanceamento, no que diz respeito ao impacto no desempenho dos modelos aqui utilizados. Essa inferência deriva do fato de que classes com grau desbalanceamento acentuado, como *amiloidose*, *esclerose* e *podocitopatia*, permitiram a construção de modelos com desempenho superior ao auferido em classes com maior representatividade, como *esclerose pura*. Nesse sentido, observou-se também que a classe mais presente no conjunto não apresentou o maior desempenho, ficando aquém de classes menos representadas. Os maiores resultados obtidos para cada classe utilizando-se modelos individuais são exibidos na Tabela 5.19.

Tabela 5.19: Resultados mais altos obtidos para cada classe com modelos individuais.

Classe	Conjunto	Classif.	F1-Score
Amiloidose	HE-amiloidose	FSL	92,60%(±3,11%)
Normal	ROS-HE-normal	CNN	96,41%(±0,62%)
Esclerose Pura	Baseline-esclerose-pura	CNN	50,42%(±1,31%)
Hiper celularidade	ROS-Curated-HE-hiper celularidade	CNN	78,79%(±1,51%)
Hiper cel. Pura	ROS-PAS-hiper celularidade-pura	CNN	27,81%(±1,39%)
Glomer. Crescente	Curated-glomerulonefrite	CNN	84,02%(±0,73%)
Nefr. Membranosa	Curated-HE-membranosa	FSL	87,73%(±3,16%)
Esclerose	PAS-esclerose	FSL	73,91%(±7,93%)
Podocitopatia	PAS-podocitopatia	FSL	75,76%(±5,76%)

Para todas as classes analisadas neste estudo, foram construídos classificadores *ensemble-based*, empregando tanto redes convolucionais convencionais quanto redes prototípicas, conforme o tipo de modelo que obteve o maior desempenho individual (maior F1-score) no respectivo conjunto de dados. Com exceção da classe *podocitopatia*, todas as demais apresentaram ganhos de desempenho com a utilização de modelos em conjunto. No entanto, observou-se que, em todos os casos, os desvios padrão associados a esses modelos foram superiores àqueles obtidos com os classificadores individuais, o que compromete a consistência dos resultados, especialmente no caso da classe *esclerose*. A avaliação da significância estatística, realizada por meio do teste de postos com sinais de Wilcoxon, indicou evidências limitadas quanto à superioridade dos modelos *ensemble-based*. A hipótese nula foi rejeitada apenas para os resultados das classes *esclerose pura* e *hiper celularidade*, enquanto para a classe *podocitopatia* foi observado o valor máximo do teste, sugerindo ausência total de diferença estatística entre os modelos comparados. Os resultados completos dos modelos *ensemble-based*, bem como os respectivos valores do teste de Wilcoxon, estão apresentados na Tabela 5.20, na qual, a coluna *Classif.* indica o tipo de modelo a ter alcançado o maior resultado, se CNN ou FSL.

Tabela 5.20: Resultados mais altos obtidos para cada classe com modelos *ensemble-based*. A coluna *Classif.* aponta o tipo de modelo a ter apresentado o maior resultado.

Classe	Conjunto	Classif.	F1-Score	Wilcoxon
Amiloidose	HE	FSL	93,73%(±3,25%)	0,625
Normal	ROS-HE	CNN	97,67%(±0,87%)	0,125
Esclerose Pura	Baseline	CNN	56,71%(±4,79%)	0,062
Hipercelularidade	ROS-Curated-HE	CNN	82,77%(±2,03%)	0,062
Hipercel. Pura	ROS-PAS	CNN	34,42%(±4,88%)	0,125
Glomer. Crescente	Curated	CNN	91,40%(±4,87%)	0,125
Nefr. Membranosa	Curated-HE	FSL	91,40%(±5,17%)	0,312
Esclerose	PAS	FSL	85,27%(±14,10%)	0,125
Podocitopatia	PAS	FSL	75,59%(±17,66%)	1,000

5.10.1 Comparação entre FSL e Métodos Comuns

Em se tratando da comparação entre modelos construídos com métodos comuns e o uso de Few-shot Learning, ficou evidenciado que FSL mostrou-se mais adequado em cenários com baixa representatividade da classe positiva, o que é consistente com os cenários de aplicação da técnica encontrados na literatura e listados nesse trabalho no Capítulo 4. O uso de CNNs convencionais, por sua vez, apresentou superioridade em cenários de maior abundância de dados. Uma possível razão para esse fenômeno pode residir no fato de que o treinamento episódico, realizado para a construção de modelos baseados em redes prototípicas, realiza uma subamostragem dos dados, não utilizando todas as amostras para a construção dos protótipos, o que pode descartar amostras com características mais relevantes, impactando negativamente esses modelos em configurações de vasta disponibilidade de dados.

Os resultados exibidos na Tabela 5.21 mostram que a superioridade das redes prototípicas foi significativa estatisticamente apenas para *amiloidose*. Já os modelos baseados em CNN foram superiores com significância estatística em três classes. Um aspecto importante a ser mencionado é o fato de que, os desvios padrão observados para os modelos FSL foram predominantemente superiores ao que se obteve com redes neurais convencionais, o que pode ser um indicativo de que redes prototípicas são menos estáveis frente à variabilidade dos dados.

Tabela 5.21: Resultados do teste de Wilcoxon.

Classe	Cenário 1	Cenário 2	valor- p	Cenário Superior
Amiloidose	CNN	FSL	0,062	FSL
Normal	CNN	FSL	0,062	CNN
Esclerose pura sem crescente	CNN	FSL	0,062	CNN
Hiper celularidade	CNN	FSL	0,312	CNN
Hiper cel. pura sem crescente	CNN	FSL	0,312	CNN
Glomerulonefrite crescente	CNN	FSL	0,062	CNN
Nefropatia membranosa	CNN	FSL	0,312	FSL
Esclerose	CNN	FSL	0,187	FSL
Podocitopatia	CNN	FSL	0,312	FSL

Na construção dos modelos baseados em FSL, a definição do tamanho do conjunto de suporte teve um papel significativo no desempenho dos modelos. Pelas observações feitas, *Support sets* grandes oferecem potencial para a introdução de ruídos que impactam negativamente a construção dos protótipos, resultando em menor desempenho. A classe mais bem representada de todo o conjunto estudado, *hiper celularidade*, apresentou classificadores com maior resultado quando treinados com um conjunto de suporte de 200 amostras, com queda no desempenho ao se realizar o treinamento com 400 e 800 amostras, como mostrado na Tabela 5.22.

Tabela 5.22: Resultados obtidos para a classe *hiper celularidade* com diferentes tamanhos para o conjunto de suporte (apenas modelos individuais). Os valores apresentados para o teste de Wilcoxon são referentes à comparação com modelos treinados com *200-shot*.

N-shot	F1-score	Wilcoxon
200-shot	78,79%($\pm 1,51\%$)	—
400-shot	76,37%($\pm 1,11\%$)	0,062
800-shot	75,29%($\pm 1,52\%$)	0,062

Fato similar foi observado para os conjuntos com maior taxa de desbalanceamento. Como mencionado no Capítulo 4, as classes *amiloidose*, *esclerose* e *podocitopatia* apresentaram modelos com maior índice de F1-score para um *support set* em que se utilizou 25% dos dados, contra a configuração de 80%, como se observa na Tabela 5.23.

Tabela 5.23: Resultados obtidos para as classes *amiloidose*, *esclerose* e *podocitopatia* com diferentes tamanhos para o conjunto de suporte (apenas modelos individuais). Os valores apresentados para o teste de Wilcoxon são referentes à comparação com modelos treinados com *(25%)-shot*.

Classe	N-shot (%)	F1-score	Wilcoxon
Amiloidose	(25%)-shot	92,67%(±0,87%)	–
Amiloidose	(80%)-shot	89,64%(±1,98%)	0,125
Esclerose	(25%)-shot	73,91%(±7,93%)	–
Esclerose	(80%)-shot	60,64%(±7,37%)	0,062
Podocitopatia	(25%)-shot	75,76%(±5,76%)	–
Podocitopatia	(80%)-shot	56,05%(±4,25%)	0,062

Um fator relevante observado nos experimentos utilizando redes prototípicas foi a superioridade dos modelos que utilizaram a *similaridade do cosseno* como função de distância, em comparação com modelos que empregaram a *distância euclidiana*. Como explicado no Capítulo 4, esse fato pode estar relacionado com a própria natureza da *similaridade do cosseno*, que a torna mais robusta à variação presente nos conjuntos de amostras glomerulares. A Tabela 5.24 apresenta uma comparação entre essas duas funções nas tarefas de classificação em que FSL apresentou o maior resultado neste trabalho.

Tabela 5.24: Comparação entre o uso da *similaridade do cosseno* ($\text{sim}_{\cos}(\mathbf{x}, \mathbf{y})$) e a *distância euclidiana* ($\|\mathbf{x} - \mathbf{y}\|_2$).

Classe	$\text{sim}_{\cos}(\mathbf{x}, \mathbf{y})$	$\ \mathbf{x} - \mathbf{y}\ _2$
Amiloidose	92,60%(±3,11%)	69,07%(±7,16%)
Nefr. Membranosa	87,73%(±3,16%)	80,16%(±3,53%)
Esclerose	73,91%(±7,93%)	60,89%(±5,73%)
Podocitopatia	75,76%(±5,76%)	63,50%(±4,27%)

5.10.2 Influência da Curadoria do Dataset

Comparando-se os resultados obtidos sobre dados refinados no processo de curadoria com o auferido sobre conjuntos não curados, observou-se que, das nove classes presentes no conjunto, o desempenho sobre dados curados obteve valores absolutos superiores em apenas três classes, a saber, *hipercelularidade*, *glomerulonefrite crescente* e *nefropatia membranosa*. No entanto, observou-se também que para as classes *amiloidose*, *normal* e *esclerose*, os resultados sobre dados curados ficou próximo ao obtido sobre dados não refinados, especialmente ao se observar os desvios padrão obtidos em cada caso. Salienta-se também que as classes menos representadas no conjunto, o que inclui *amiloidose* e *esclerose*, não sofreram o acréscimo de novas amostras no processo de curadoria, o que contribuiu para o aumento do desbalançamento nesses casos, o que pode ter contribuído para o que fora observado nos experimentos.

5.10.3 Influência dos Corantes

Em se tratando dos resultados obtidos sobre conjuntos restritos a um determinado corante, observou-se a predominância no uso desses conjuntos em relação à utilização de conjuntos multi-corante. Das nove classes do conjunto, sete apresentaram desempenho superior com modelos treinados sobre conjuntos restritos, com quatro desses resultados sendo obtidos em amostras tingidas com HE e três em amostras coloridas com PAS. Salienta-se ainda o fato de que as classes que apresentaram maiores resultados com valores absolutos com conjuntos multi-corante, especificamente *esclerose pura sem crescente* e *glomerulonefrite crescente* tiveram resultados próximos com conjuntos segmentados, com pontuações respectivas no teste de Wilcoxon de 0,125 e 0,812, indicando pouca relevância estatística na superioridade dos modelos treinados em contexto multicorante. Essa observação pode ser entendida como um indício de que conjuntos segmentados apresentem menor complexidade, permitindo a extração de melhores *features* para a construção dos classificadores.

A Tabela 5.25 apresenta os resultados do teste de Wilcoxon para a comparação entre os desempenhos obtidos com conjuntos multicorante e monocorante. Os valores reportados evidenciam a significância estatística da superioridade dos modelos treinados com dados segmentados por corante. Conforme já discutido no parágrafo anterior, nos dois únicos cenários em que os conjuntos multicorante obtiveram desempenho superior, não houve rejeição da hipótese nula. Por outro lado, os resultados do teste aplicados aos cenários em que os conjuntos monocorante se destacaram indicam que, dentre as sete classes analisadas, cinco apresentaram diferenças estatisticamente significativas, reforçando a eficácia da segmentação por corante na melhoria do desempenho dos modelos.

Tabela 5.25: Resultados do teste de Wilcoxon. Comparação entre modelos treinados com conjuntos monocorante (Mono) e multicorante (Multi).

Classe	Multi	Mono	valor- <i>p</i>	Cenário Superior
Amiloidose	Baseline	HE	0,062	Mono
Normal	ROS-Curated	ROS-HE	0,062	Mono
Esclerose pura	Baseline	HE	0,125	Multi
Hipercelularidade	ROS-Curated	ROS-Curated-HE	0,062	Mono
Hipercel. pura	Baseline	ROS-PAS	0,125	Mono
Glomer. crescente	Curated	Curated-HE	0,812	Multi
Nefr. membranosa	ROS-Curated	Curated-HE	0,062	Mono
Esclerose	Baseline	PAS	0,125	Mono
Podocitopatia	Baseline	PAS	0,062	Mono

5.10.4 Particularidades das Classes Esclerose Pura e Hipercelularidade Pura

As classes *esclerose pura sem crescente* e *hipercelularidade pura sem crescente* apresentaram as métricas mais baixas dentre as nove classes aqui estudadas. Como

explicado no Capítulo 2, essas classes têm uma relação direta com suas contrapartes esclerose e hiper celularidade, no sentido de que suas amostras apresentam as mesmas respectivas lesões, porém, sem a presença de outras patologias.

Analisando-se as predições incorretas, observou-se que, para a classe *hipercelularidade pura sem crescente*, o maior índice de falsos positivos incorreu na classe *hipercelularidade*. No entanto, não foi observado comportamento similar para a classe *esclerose pura sem crescente*, para a qual o maior parte dos erros do tipo *I* também ocorreu para a lesão *hipercelularidade*. Em ambos os casos, pode ter havido a influência da alta representatividade da classe *hipercelularidade*, uma vez que esta foi a lesão com maior número de amostras nos conjuntos utilizados.

5.10.5 Comportamento das Métricas Precisão e Revocação (Precision e Recall)

Em todos os resultados apresentados neste capítulo, observou-se uma predominância dos valores de *recall* em relação à *precision*, o que indica uma maior incidência de falsos positivos nos experimentos realizados. A maior discrepância entre essas métricas foi registrada para a classe *hipercelularidade pura sem crescente*, cuja média de *recall* foi de aproximadamente 68%, enquanto a *precision* permaneceu em torno de 16%. Tal comportamento pode estar associado à estratégia de aprendizado sensível ao custo adotada nos experimentos, a qual impôs penalizações mais severas à ocorrência de falsos negativos, com o intuito de mitigar os efeitos do desbalanceamento entre as classes.

5.10.6 Complexidade dos Conjuntos

As métricas obtidas na análise de complexidade dos conjuntos estudados, evidenciam a heterogeneidade dos cenários encontrados neste trabalho. A análise sugere que as classes *normal* e *amiloidose* apresentam conjuntos de menor complexidade, o que é enfatizado especialmente pelos seus resultados para a métrica de linearidade. Por outro lado, classes como *esclerose pura sem crescente* e *hipercelularidade pura sem crescente* apresentaram métricas que indicam uma complexidade acentuada, com destaque para o que foi obtido para *N1*. Essas observações podem indicar a necessidade de outras abordagens para lidar com problemas de classificação mais complexos, além do que foi apresentado neste texto.

Este capítulo apresentou os resultados obtidos em todos os experimentos conduzidos neste trabalho. O próximo capítulo é dedicado à análise crítica desses resultados, incluindo discussões, interpretações e considerações relevantes.

Capítulo 6

Conclusões

Este trabalho teve por objetivo principal desenvolver, otimizar e avaliar sistematicamente classificadores automáticos para a classificação de lesões em glomérulos renais, enfrentando o cenário imposto pelo severo desbalanceamento de classes e pela complexidade intrínseca dos dados histopatológicos. Foi utilizado um conjunto de dados composto por amostras de nove classes, coradas em quatro diferentes corantes. Nos experimentos realizados, foram construídos classificadores binários, tomando-se cada uma das classes presentes no conjunto como sendo a classe positiva, apresentando assim, diferentes cenários de desbalanceamento.

Os classificadores construídos alcançaram, em sua maioria, taxas de F1-score acima dos 70%. Das nove classes presentes no conjunto estudado, apenas duas tiveram desempenho abaixo desse patamar. Ressalta-se o resultado obtido para *amiloidose*, uma classe com taxa de desbalanceamento de aproximadamente 1 : 30 e sobre a qual se obteve 92,60%, chegando a 93,73% com um classificador *ensemble-based*. Outro resultado importante é o observado para a classe *normal* sobre a qual todos os modelos construídos alcançaram altas taxas de F1-score, chegando a 97,67%. Como será melhor descrito a seguir, FSL apresentou maior desempenho em cenários de relativa escassez de dados e houve superioridade dos modelos treinados em cenários mono-corante.

Este trabalho teve como um de seus objetivos específicos, analisar comparativamente a eficácia de abordagens baseadas em *Few-Shot Learning* em redes neurais convolucionais convencionais para o tratamento do desbalanceamento severo de classes, identificando os cenários mais adequados para a aplicação de cada técnica ou suas combinações. Para isso, foram desenvolvidos classificadores baseados em CNN e FSL, que tiveram o seu desempenho avaliado e comparado para cada configuração do conjunto de dados aqui utilizado.

Na comparação entre técnicas convencionais e o uso de FSL, observou-se que esta última, especialmente quando implementada com redes prototípicas, demonstra maior adequação a cenários marcados pela escassez de amostras. Isso se deve à própria

lógica dessas redes, que se baseiam na seleção de um número limitado de exemplos para a construção dos protótipos. Em contextos com abundância de dados, essa estratégia pode resultar em uma subamostragem implícita que descarta informações relevantes, comprometendo o desempenho dos modelos. Assim, a eficácia das redes prototípicas tende a ser reduzida em cenários onde há uma maior disponibilidade de dados.

Um dos objetivos específicos foi também investigar o impacto da qualidade dos dados no desempenho dos modelos de ML aqui utilizados. Para isso, conjunto original foi submetido a um processo de curadoria manual, no qual se procurou mitigar a existência de possíveis ruídos que trariam impactos negativos à generalização dos classificadores.

O processo de curadoria resultou em ganho de desempenho para três das classes analisadas neste estudo. No entanto, é importante destacar que o refinamento do conjunto não impactou de forma significativa as classes menos representadas, ocasionando, inclusive, um aumento na taxa de desbalanceamento para essas classes. Além disso, nos casos em que as maiores taxas de F1-score foram obtidas com dados não curados, o desempenho alcançado com os dados refinados mostrou-se muito próximo, não permitindo inferir, de forma cabal, se a curadoria foi um fator negativo nesses cenários.

O trabalho aqui apresentado também se propôs a avaliar a influência da segmentação de dados por tipo de corante histológico (multi-corante versus mono-corante). Para este fim, foram utilizados neste estudo conjuntos segmentados pelo corante HE, bem como aqueles restritos à coloração PAS. O objetivo desses experimentos foi observar se a restrição por corante traria algum ganho de desempenho pela redução da complexidade dos conjuntos e potencial mitigação de sub-conceitos.

Os experimentos realizados com amostras restritas a um único corante apresentaram, de modo geral, desempenho superior àqueles conduzidos sobre conjuntos com múltiplos corantes, com destaque para os conjuntos exclusivos da coloração HE. Esse ganho de desempenho pode estar associado à redução da complexidade dos dados, especialmente pela mitigação de sub-conceitos introduzidos pela variabilidade entre diferentes corantes. Tal evidência sugere que a construção de modelos mono-corante pode ser uma estratégia mais eficaz no contexto da classificação de lesões glomerulares.

Este trabalho ainda avaliou a complexidade de cada conjunto obtido a partir do *dataset* original, utilizando abordagens consolidadas na literatura, como a pontuação de silhueta, a métrica N1 e a separabilidade linear. Os resultados obtidos pelos modelos, quando analisados em conjunto com as métricas de complexidade dos dados, indicam que o desbalanceamento não é o único fator determinante no processo de aprendizado dos classificadores. A complexidade dos dados mostrou-se um aspecto relevante — e, em alguns casos, possivelmente decisivo — para o desempenho final dos modelos produzidos.

Os resultados obtidos para as classes *esclerose pura sem crescente* (50,42%) e *hipercelularidade pura sem crescente* (27,81%) indicam a necessidade de uma investigação mais aprofundada dessas categorias de forma individualizada, a fim de identificar outros fatores que possam ter contribuído para o baixo desempenho observado. Uma possível abordagem para mitigar esse problema seria a implementação de uma arquitetura de classificação em cadeia, na qual uma etapa inicial realizaria uma classificação preliminar para detectar a presença de *esclerose* ou *hipercelularidade*, seguida por uma etapa subsequente responsável por refinar a predição, distinguindo entre *esclerose pura sem crescente* e *hipercelularidade pura sem crescente*.

Este documento apresentou um estudo sobre a classificação de lesões glomerulares a partir de imagens de biópsias renais, considerando cenários marcados por desbalançamento de dados. Até o momento, não se tem conhecimento, na literatura acadêmica, de trabalhos que tenham abordado a classificação de lesões glomerulares com a abrangência aqui apresentada. Espera-se que os achados relatados neste estudo possam servir como referência para investigações futuras na área da nefropatologia, especialmente no que se refere ao desenvolvimento de ferramentas computacionais de apoio ao diagnóstico dessas lesões.

6.1 Trabalhos Futuros

Considerando-se o conjunto de dados utilizado neste estudo e os resultados obtidos pelos diferentes modelos para cada uma das classes analisadas, pode-se inferir a necessidade de novas estratégias tanto voltadas para os dados quanto relacionadas aos modelos de ML a serem empregados na solução aqui proposta.

No tocante ao conjunto de dados, o processo de curadoria realizado ainda mostra-se inconclusivo para algumas classes. Salienta-se que esse processo não pôde ser realizado de forma completa, uma vez que as amostras de glomérulos parcialmente representados não puderam ser removidas do conjunto, sob pena de inviabilizar a aferição dos modelos produzidos por meio de validação cruzada sobre os conjuntos de dados das classes menos representadas. Portanto, um incremento do conjunto de dados se faz necessário para uma melhor avaliação do impacto da qualidade das amostras no desempenho dos modelos. Considerando o elevado custo operacional associado à coleta e anotação de amostras glomerulares, recomenda-se o desenvolvimento de um segmentador automático, com o objetivo de reduzir, ainda que parcialmente, a dependência de profissionais da patologia nesse processo.

Com base nos resultados obtidos — em especial no desempenho alcançado para a classe *normal* — entende-se que a adoção de uma estratégia de classificação sequencial, na qual se verifica inicialmente a presença ou ausência de lesão, pode se mostrar mais eficaz para a classificação automática de lesões glomerulares do que a utilização de classificadores paralelos. Nessa abordagem, cada amostra seria primeiramente avaliada por um classificador treinado para identificar padrões normais,

sendo encaminhada aos demais modelos apenas em caso de resultado negativo nessa primeira etapa.

Outro fator relevante a ser considerado é a possibilidade de coexistência de múltiplas patologias em uma mesma biópsia glomerular, o que implica que uma única amostra pode conter mais de uma lesão. Diante disso, recomenda-se que futuros desenvolvimentos contemplem a construção de um classificador *multi-label*, capaz de identificar simultaneamente todas as lesões presentes em cada imagem. Tal implementação poderia ainda ser derivada da combinação dos classificadores já desenvolvidos neste trabalho, considerando-se os respectivos níveis de confiança apresentados por cada modelo.

No que se refere às técnicas e algoritmos, salienta-se que, neste estudo, a construção de modelos com a abordagem *few-shot learning* limitou-se à utilização de redes prototípicas, permanecendo em aberto a exploração de outras abordagens da literatura, como as *redes siamesas* e as *matching networks*. Além do escopo de FSL, técnicas mais recentes de *machine learning* têm apontado para o uso de aprendizado autossupervisionado (*self-supervised learning*) (Rani et al., 2023), *visual transformers* (ViT) (Lu et al., 2024) e *foundational models* (Chen et al., 2024a) na classificação e rotulagem de imagens. O uso de aprendizado autossupervisionado pode apresentar uma solução para problema de rotulação de amostras, especialmente se combinado com o segmentador também proposto neste capítulo. O uso ViT apresenta potencial para ganhos desempenho devido à capacidade desses modelos em capturar relações globais contextuais das imagens analisadas. Já o uso de *foundational models* pode oferecer uma ferramenta de ML capaz capturar o fundamento da classificação e análise glomerular, podendo adaptar esse conhecimento a outros contextos e tarefas. Com base no que tem se observado na literatura, é plausível que a adoção dessas metodologias traga ganhos em relação aos resultados obtidos neste trabalho, sobretudo considerando-se a possibilidade de ampliação do conjunto de dados, uma vez que tais técnicas, especialmente no tocante a *foundational models*, apresentam uma demanda substancial por grandes volumes de dados.

Considerando-se todos os resultados obtidos, é claramente observável que os modelos produzidos para as classes *esclerose pura sem crescente* e *hipercelularidade pura sem crescente* apresentaram desempenho consideravelmente inferior em comparação às demais classes analisadas neste estudo. Esse fato indica a necessidade de uma investigação mais aprofundada e individualizada dessas classes, com o objetivo de identificar possíveis fatores adicionais que justifiquem o baixo desempenho alcançado.

Referências

- Abbas, Q. (2023). An intelligent medical image classification system using few-shot learning. *Concurrency and Computation: Practice and Experience*, 35(2):e7451.
- Abels, E., Pantanowitz, L., Aeffner, F., Zarella, M. D., van der Laak, J., Bui, M. M., Vemuri, V. N., Parwani, A. V., Gibbs, J., Agosto-Arroyo, E., Beck, A. H., e Kozlowski, C. (2019). Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the digital pathology association. *The Journal of Pathology*, 249(3):286–294.
- Al Mushafi, A., Ooi, J. D., e Odobasic, D. (2021). Crescentic glomerulonephritis: Pathogenesis and therapeutic potential of human amniotic stem cells. *Front. Physiol.*, 12:724186.
- Barbosa, D. K., Matteucci, D. R., e Mesquita, P. G. M. (2015). Amiloidose renal em paciente portadora de hiv: Relato de caso/ renal amyloidosis in a patient with hiv: Case report. *Revista Ciências em Saúde*, 4(4):73–79.
- Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In *Lecture Notes in Computer Science*, Lecture notes in computer science, páginas 437–478. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Billion Polak, P., Prusa, J. D., e Khoshgoftaar, T. M. (2024). Low-shot learning and class imbalance: a survey. *J. Big Data*, 11(1).
- Calumby, R. T., Duarte, A. A., Angelo, M. F., Santos, E., Sarder, P., dos Santos, W. L. C., e Oliveira, L. R. (2023). Toward real-world computational nephropathology. *Clin. J. Am. Soc. Nephrol.*, 18(6):809–812.
- Chagas, P., Souza, L., Araújo, I., Aldeman, N., Duarte, A., Angelo, M., Dos-Santos, W. L. C., e Oliveira, L. (2020). Classification of glomerular hypercellularity using convolutional features and support vector machine. *Artif. Intell. Med.*, 103(101808):101808.
- Chen, R. J., Ding, T., Lu, M. Y., Williamson, D. F. K., Jaume, G., Song, A. H., Chen, B., Zhang, A., Shao, D., Shaban, M., Williams, M., Oldenburg, L., Weishaupt, L. L., Wang, J. J., Vaidya, A., Le, L. P., Gerber, G., Sahai, S., Williams, W., e Mahmood, F. (2024a). Towards a general-purpose foundation model for computational pathology. *Nat. Med.*, 30(3):850–862.

- Chen, T., Zhang, D., e Lee, D.-J. (2024b). A new joint training method for facial expression recognition with inconsistently annotated and imbalanced data. *Electronics*, 13(19).
- Cheon, M. (2024). Kolmogorov-Arnold network for satellite image classification in remote sensing.
- Chicco, D., Warrens, M. J., e Jurman, G. (2021). The matthews correlation coefficient (mcc) is more informative than cohen's kappa and brier score in binary classification assessment. *IEEE Access*, 9:78368–78381.
- Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., e Prior, F. (2013). The cancer imaging archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging*, 26(6):1045–1057.
- Cortes, D., Bermejo, B., e Juiz, C. (2024). The use of CNNs in VR/AR/MR/XR: a systematic literature review. *Virtual Real.*, 28(3).
- Cui, M. e Zhang, D. Y. (2021). Artificial intelligence and computational pathology. *Laboratory Investigation*, 101(4):412–422.
- Dai, Q., wei Liu, J., e hui Shi, Y. (2023). Class-overlap undersampling based on schur decomposition for class-imbalance problems. *Expert Systems with Applications*, 221:119735.
- Dai, W., Li, D., Tang, D., Wang, H., e Peng, Y. (2022). Deep learning approach for defective spot welds classification using small and class-imbalanced datasets. *Neurocomputing*, 477:46–60.
- Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.*, 7:1–30.
- Deng, Y. e Li, H. (2023). Deep learning for few-shot white blood cell image classification and feature learning. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 11(6):2081–2091.
- Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, 10(7):1895–1923.
- Dio, L. J. A. D. (1999). *Tratado de anatomia aplicada*. Pólus Editorial, São Paulo, 1 edição.
- Duarte, D., Andriolo, U., e Gonçalves, G. (2020). Addressing the class imbalance problem in the automatic image classification of coastal litter from orthophotos derived from uas imagery. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.*, V-3-2020:439–445.

- Duarte, H. E. (2014). *Anatomia Humana*. Universidade Federal de Santa Catarina, primeira edição.
- Dumitru, Ian Goodfellow, W. C. Y. B. (2013). Challenges in representation learning: Facial expression recognition challenge.
- Edward, J., Rosli, M. M., e Seman, A. (2023). A new multi-class rebalancing framework for imbalance medical data. *IEEE Access*, 11:92857–92874.
- Erickson, B. J. e Kitamura, F. (2021). Magician’s corner: 9. performance metrics for machine learning models. *Radiol. Artif. Intell.*, 3(3):e200126.
- Escobar Díaz Guerrero, R., Carvalho, L., Bocklitz, T., Popp, J., e Oliveira, J. L. (2024). A data augmentation methodology to reduce the class imbalance in histopathology images. *J Imaging Inform Med*, 37(4):1767–1782.
- Fei-Fei, L., Fergus, R., e Perona, P. (2006). One-shot learning of object categories. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(4):594–611.
- Fernandez, A., Garcia, S., Galar, M., Prati, R. C., Krawczyk, B., e Herrera, F. (2018). *Learning from imbalanced data sets*. Springer International Publishing, Cham, Switzerland, 1 edição.
- Fink, M. (2004). Object classification from a single example utilizing class relevance metrics. In Saul, L., Weiss, Y., e Bottou, L., editores, *Advances in Neural Information Processing Systems*, volume 17. MIT Press.
- Gardner, E. et al. (1988). *Anatomia*. Guanabara Koogan, Rio de Janeiro, 4 edição.
- Gidaris, S. e Komodakis, N. (2018). Dynamic few-shot visual learning without forgetting. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, páginas 4367–4375.
- Goldenberg, R. C. d. S. e Carvalho, A. C. C. d., editores (2012). *Resident stem cells and regenerative therapy*. Academic Press, San Diego, CA.
- Goodfellow, I., Bengio, Y., e Courville, A. (2016). *Deep Learning*. MIT Press. <http://www.deeplearningbook.org>.
- Guan, J., Liu, J., Sun, J., Feng, P., Shuai, T., e Wang, W. (2020). Meta metric learning for highly imbalanced aerial scene classification. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, páginas 4047–4051.
- Gurina, T. S. e Simms, L. (2023). Histology, staining. StatPearls [Internet]. Updated 2023 May 1. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK557663/>.

- Haque, M. A., Marwaha, S., Deb, C. K., Nigam, S., Arora, A., Hooda, K. S., Soujanya, P. L., Aggarwal, S. K., Lall, B., Kumar, M., Islam, S., Panwar, M., Kumar, P., e Agrawal, R. C. (2022). Deep learning-based approach for identification of diseases of maize crop. *Sci. Rep.*, 12(1):6334.
- He, K., Zhang, X., Ren, S., e Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.
- Hosseini, M. S., Bejnordi, B. E., Trinh, V. Q.-H., Chan, L., Hasan, D., Li, X., Yang, S., Kim, T., Zhang, H., Wu, T., Chinniah, K., Maghsoudlou, S., Zhang, R., Zhu, J., Khaki, S., Buin, A., Chaji, F., Salehi, A., Nguyen, B. N., Samaras, D., e Plataniotis, K. N. (2024). Computational pathology: A survey review and the way forward. *Journal of Pathology Informatics*, 15:100357.
- Ilmawati, F. I., Kusrini, K., e Hidayat, T. (2024). Optimizing facial expression recognition with image augmentation techniques: VGG19 approach on FERC dataset. *Sinkron*, 8(2):632–640.
- Johnson, J. M. e Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. *J. Big Data*, 6(1).
- Kang Wang, X. W. e Cheng, Y. (2022). Few-shot aerial image classification with deep economic network and teacher knowledge. *International Journal of Remote Sensing*, 43(13):5075–5099.
- Khaleel, M., Idris, A., Tavanapong, W., Pratt, J. R., Oh, J., e C. de Groen, P. (2023). Visactive: Visual-concept-based active learning for image classification under class imbalance. *ACM Trans. Multimedia Comput. Commun. Appl.*, 20(3).
- Koch, G. R. (2015). Siamese neural networks for one-shot image recognition.
- Kopp, J. B., Anders, H.-J., Susztak, K., Podestà, M. A., Remuzzi, G., Hildebrandt, F., e Romagnani, P. (2020). Podocytopathies. *Nat. Rev. Dis. Primers*, 6(1):68.
- Liu, Z., Wang, R., Yang, J., e Zhang, W. (2022). Research on fine-tuning cnn for cancer diagnosis with gene expression data. In *Proceedings of the 2022 14th International Conference on Machine Learning and Computing, ICMLC '22*, página 140–145, New York, NY, USA. Association for Computing Machinery.
- Lorena, A. C., Garcia, L. P. F., Lehmann, J., Souto, M. C. P., e Ho, T. K. (2019). How complex is your classification problem? a survey on measuring classification complexity. *ACM Comput. Surv.*, 52(5).
- Lu, M. Y., Chen, B., Williamson, D. F. K., Chen, R. J., Liang, I., Ding, T., Jaume, G., Odintsov, I., Le, L. P., Gerber, G., Parwani, A. V., Zhang, A., e Mahmood, F. (2024). A visual-language foundation model for computational pathology. *Nat. Med.*, 30(3):863–874.

- Mahbub, T., Obeid, A., Javed, S., Dias, J., Hassan, T., e Werghi, N. (2024). Center-focused affinity loss for class imbalance histology image classification. *IEEE Journal of Biomedical and Health Informatics*, 28(2):952–963.
- Maier-Hein, L., Reinke, A., Godau, P., Tizabi, M. D., Buettner, F., Christodoulou, E., Glocker, B., Isensee, F., Kleesiek, J., Kozubek, M., Reyes, M., Riegler, M. A., Wiesenfarth, M., Kavur, A. E., Sudre, C. H., Baumgartner, M., Eisenmann, M., Heckmann-Nötzl, D., Rädtsch, T., Acion, L., Antonelli, M., Arbel, T., Bakas, S., Benis, A., Blaschko, M. B., Cardoso, M. J., Cheplygina, V., Cimini, B. A., Collins, G. S., Farahani, K., Ferrer, L., Galdran, A., van Ginneken, B., Haase, R., Hashimoto, D. A., Hoffman, M. M., Huisman, M., Jannin, P., Kahn, C. E., Kainmueller, D., Kainz, B., Karargyris, A., Karthikesalingam, A., Kofler, F., Kopp-Schneider, A., Kreshuk, A., Kurc, T., Landman, B. A., Litjens, G., Madani, A., Maier-Hein, K., Martel, A. L., Mattson, P., Meijering, E., Menze, B., Moons, K. G. M., Müller, H., Nichyporuk, B., Nickel, F., Petersen, J., Rajpoot, N., Rieke, N., Saez-Rodriguez, J., Sánchez, C. I., Shetty, S., van Smeden, M., Summers, R. M., Taha, A. A., Tiulpin, A., Tsaftaris, S. A., Van Calster, B., Varoquaux, G., e Jäger, P. F. (2024). Metrics reloaded: recommendations for image analysis validation. *Nat. Methods*, 21(2):195–212.
- Masters, D. e Luschi, C. (2018). Revisiting small batch training for deep neural networks.
- McDonald, J. H. (2014). *Handbook of Biological Statistics*. Sparky House Publishing, Baltimore, Maryland, 3rd edição.
- Medela, A., Picon, A., Saratxaga, C. L., Belar, O., Cabezón, V., Cicchi, R., Bilbao, R., e Glover, B. (2019). Few shot learning in histopathological images:reducing the need of labeled data on biological datasets. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, páginas 1860–1864.
- Megahed, F. M., Chen, Y.-J., Megahed, A., Ong, Y., Altman, N., e Krzywinski, M. (2021). The class imbalance problem. *Nature Methods*, 18(11):1270–1272.
- Mollahosseini, A., Hasani, B., e Mahoor, M. H. (2019). AffectNet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Trans. Affect. Comput.*, 10(1):18–31.
- Montanari, T. (2016). *Histologia: texto, atlas e roteiro de aulas práticas*. Ed. da autora, Porto Alegre, 3 edição.
- Monteiro, N. F. e Diz, M. C. E. (2015). Difficulties in the diagnosis of primary amyloidosis: case report. *Revista Médica de Minas Gerais*, 25(2):268–274.
- Moore, K. L. e Dalley, A. F. (2007). *Anatomia orientada para a clínica*. Guanabara Koogan, Rio de Janeiro, 5 edição.

- Moraes, C. e Colicigno, P. (2007). Estudo morfofuncional do sistema renal. *Ensaios e Ciência C Biológicas Agrárias e da Saúde*, 1:161–167.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press.
- National Institute of Diabetes and Digestive and Kidney Diseases (U.S.) (2022). Homepage. Accessed: 27 September 2024.
- Nourivandi, T., Hinduja, S., Srivastava, S., Cohn, J. F., e Canavan, S. (2024). Mitigating class imbalance for facial expression recognition using SMOTE on deep features. In *2024 IEEE 18th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE.
- Ochal, M., Patacchiola, M., Vazquez, J., Storkey, A., e Wang, S. (2023). Few-shot learning with class imbalance. *IEEE Transactions on Artificial Intelligence*, 4(5):1348–1358.
- Oliveira, A. d. A. e Neto, F. H. C. (2015). *Anatomia e Fisiologia: a incrível máquina do corpo humano*. Editora UECE. https://educapes.capes.gov.br/bitstream/capes/432728/2/Livro_Anatomia%20e%20Fisiologia%20Humana.PDF.
- Orriols-Puig, A., Macià, N., e Ho, T. (2010). Dcol: Data complexity library in c++ (documentation).
- Parnami, A. e Lee, M. (2022). Learning from few examples: A summary of approaches to few-shot learning.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., e Chintala, S. (2019). *PyTorch: an imperative style, high-performance deep learning library*. Curran Associates Inc., Red Hook, NY, USA.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., e Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.-T., Lux, M., Schmidt, P. T., Riegler, M., e Halvorsen, P. (2017). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection.

- Quan, H., Li, X., Hu, D., Nan, T., e Cui, X. (2024). Dual-channel prototype network for few-shot pathology image classification. *IEEE Journal of Biomedical and Health Informatics*, 28(7):4132–4144.
- Raj, R., Mathew, J., Kannath, S. K., e Rajan, J. (2022). Crossover based technique for data augmentation. *Computer Methods and Programs in Biomedicine*, 218:106716.
- Rani, V., Nabi, S. T., Kumar, M., Mittal, A., e Kumar, K. (2023). Self-supervised learning: A succinct review. *Arch. Comput. Methods Eng.*, 30(4):2761–2775.
- Reiser, J. e Altintas, M. M. (2016). Podocytes. *F1000Res.*, 5:114.
- Ronco, P., Beck, L., Debiec, H., Fervenza, F. C., Hou, F. F., Jha, V., Sethi, S., Tong, A., Vivarelli, M., e Wetzels, J. (2021). Membranous nephropathy. *Nat. Rev. Dis. Primers*, 7(1):69.
- Sakri, S. e Basheer, S. (2023). Fusion model for classification performance optimization in a highly imbalance breast cancer dataset. *Electronics*, 12(5).
- Simonyan, K. e Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556.
- Snell, J., Swersky, K., e Zemel, R. (2017a). Prototypical networks for few-shot learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS’17, página 4080–4090, Red Hook, NY, USA. Curran Associates Inc.
- Snell, J., Swersky, K., e Zemel, R. (2017b). Prototypical networks for few-shot learning. In Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., e Garnett, R., editores, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.
- Song, B., Li, S., Sunny, S., Gurushanth, K., Mendonca, P., Mukhia, N., Patrick, S., Gurudath, S., Raghavan, S., Tsusennaro, I., Leivon, S. T., Kolur, T., Shetty, V., Bushan, V., Ramesh, R., Peterson, T., Pillai, V., Wilder-Smith, P., Sigamani, A., Suresh, A., Kuriakose, M. A., Birur, P., e Liang, R. (2021). Classification of imbalanced oral cancer image data from high-risk population. *J. Biomed. Opt.*, 26(10).
- Starczewski, A. e Krzyżak, A. (2015). Performance evaluation of the silhouette index. In Rutkowski, L., Korytkowski, M., Scherer, R., Tadeusiewicz, R., Zadeh, L. A., e Zurada, J. M., editores, *Artificial Intelligence and Soft Computing*, páginas 49–58, Cham. Springer International Publishing.
- Sümer, Ö., Hellmann, F., Hustinx, A., Hsieh, T.-C., André, E., e Krawitz, P. (2023a). Few-shot meta-learning for recognizing facial phenotypes of genetic disorders. *Stud. Health Technol. Inform.*, 302:932–936.

- Sümer, Ö., Hellmann, F., Hustinx, A., Hsieh, T.-C., André, E., e Krawitz, P. (2023b). Few-shot meta-learning for recognizing facial phenotypes of genetic disorders. *Stud. Health Technol. Inform.*, 302:932–936.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., e Wojna, Z. (2015). Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567.
- Tan, M. e Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. *CoRR*, abs/1905.11946.
- Timm, L. d. L. (2005). Técnicas rotineiras de preparação e análise de lâminas histológicas. *Caderno La Salle XI*, 2(1):231–239.
- Titoriya, A. K. e Singh, M. P. (2022). Few-shot learning on histopathology image classification. In *2022 International Conference on Computational Science and Computational Intelligence (CSCI)*, páginas 251–256.
- Toennies, K. D. (2024). *An Introduction to Image Classification: From Designed Models to End-to-End Learning*. Springer Nature Singapore, Singapore.
- Torralba, A., Isola, P., e Freeman, W. T. (2024). *Foundations of Computer Vision*. MIT Press.
- Tremblay, J., Prakash, A., Acuna, D., Brophy, M., Jampani, V., Anil, C., To, T., Cameracci, E., Boochoon, S., e Birchfield, S. (2018). Training deep networks with synthetic data: Bridging the reality gap by domain randomization. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, páginas 1082–1090, Salt Lake City, UT, USA.
- Universidade Estadual Paulista (2024). Como a urina é formada? https://www2.ibb.unesp.br/Museu_Escola/2_qualidade_vida_humana/Museu2_qualidade_corpo_renal3.htm. Accessed: 25 September 2024.
- University of Utah (2022). Kidney anatomy. <https://webpath.med.utah.edu/RENAHTML/RENAL116.html>. Accessed: 25 September 2024.
- Vinyals, O., Blundell, C., Lillicrap, T., Kavukcuoglu, K., e Wierstra, D. (2017). Matching networks for one shot learning. In *Advances in Neural Information Processing Systems (NeurIPS)*.
- Vinyals, O., Blundell, C., Lillicrap, T. P., Kavukcuoglu, K., e Wierstra, D. (2016). Matching networks for one shot learning. *CoRR*, abs/1606.04080.
- Walsh, R. e Tardy, M. (2023). A comparison of techniques for class imbalance in deep learning classification of breast cancer. *Diagnostics*, 13(1).
- Wang, Y., Yao, Q., Kwok, J. T., e Ni, L. M. (2020). Generalizing from a few examples: A survey on few-shot learning. *ACM Comput. Surv.*, 53(3).

- Wang, Y.-C. e Cheng, C.-H. (2021). A multiple combined method for rebalancing medical data with class imbalances. *Computers in Biology and Medicine*, 134:104527.
- Wen Xie, T. L. e Sun, H. (2024). Channel adaptive cvfcn using a new transfer method for polsar terrain classification. *International Journal of Remote Sensing*, 45(18):6521–6543.
- Whittier, W. L. e Korbet, S. M. (2004). Renal biopsy: update. *Curr. Opin. Nephrol. Hypertens.*, 13(6):661–665.
- Wold, S., Esbensen, K., e Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2(1):37–52. Proceedings of the Multivariate Statistical Workshop for Geologists and Geochemists.
- Wu, J. (2017). Introduction to convolutional neural networks. *National Key Lab for Novel Software Technology*.
- Xie, Z., Zhang, Z., Cao, Y., Lin, Y., Bao, J., Yao, Z., Dai, Q., e Hu, H. (2022). Simmim: a simple framework for masked image modeling. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 9643–9653.
- Xu, M., Yoon, S., Fuentes, A., e Park, D. S. (2023). A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, 137:109347.
- Yamashita, R., Nishio, M., Do, R. K. G., e Togashi, K. (2018). Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9(4):611–629.
- Zhang, R. e Liu, Q. (2022). Learning with few samples in deep learning for image classification, a mini-review. *Front. Comput. Neurosci.*, 16:1075294.
- Zhu, L. e Yang, Y. (2020). Inflated episodic memory with region self-attention for long-tailed visual recognition. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, páginas 4343–4352.

Apêndice A

Descrição dos Conjuntos de Dados Utilizados nos Experimentos

Tabela A.1: Descrição dos conjuntos de dados criados para este trabalho.

Conjunto	Classes de interesse	Detalhes
Baseline-amiloidose	Amiloidose	Conjuntos contendo todas as classes e corantes, sem rebalanceamento.
Baseline-normal	Normal	
Baseline-esclerose-pura	Esclerose pura sem crescente	
Baseline-hipercelularidade	Hipercelularidade	
Baseline-hipercelularidade-pura	Hipercelularidade pura sem crescente	
Baseline-glomerulonefrite	Glomerulonefrite crescente	
Baseline-membranosa	Nefropatia membranosa	
Baseline-esclerose	Esclerose	
Baseline-podocitopatia	Podocitopatia	
ROS-amiloidose	Amiloidose	Conjuntos contendo todas as classes e corantes, gerados com a estratégia de rebalanceamento ROS.
ROS-normal	Normal	
ROS-esclerose-pura	Esclerose pura sem crescente	
ROS-hipercelularidade	Hipercelularidade	
ROS-hipercelularidade-pura	Hipercelularidade pura sem crescente	
ROS-glomerulonefrite	Glomerulonefrite crescente	
ROS-membranosa	Nefropatia membranosa	
Continua na próxima página		

Tabela A.1 – continuação da página anterior

Conjunto	Classes de interesse	Detalhes
ROS-esclerose	Esclerose	Conjuntos restritos a amostras tingidas com o corante HE, sem rebalanceamento.
ROS-podocitopatia	Podocitopatia	
HE-amiloidose	Amiloidose	
HE-normal	Normal	
HE-esclerose-pura	Esclerose pura sem crescente	
HE-hipercelularidade	Hipercelularidade	
HE-hipercelularidade-pura	Hipercelularidade pura sem crescente	
HE-glomerulonefrite	Glomerulonefrite crescente	
HE-membranosa	Nefropatia membranosa	
HE-esclerose	Esclerose	
HE-podocitopatia	Podocitopatia	Conjuntos restritos a amostras tingidas com o corante HE, gerados com rebalanceamento ROS.
ROS-HE-amiloidose	Amiloidose	
ROS-HE-normal	Normal	
ROS-HE-esclerose-pura	Esclerose pura sem crescente	
ROS-HE-hipercelularidade	Hipercelularidade	
ROS-HE-hipercelularidade-pura	Hipercelularidade pura sem crescente	
ROS-HE-glomerulonefrite	Glomerulonefrite crescente	
ROS-HE-membranosa	Nefropatia membranosa	
ROS-HE-esclerose	Esclerose	
ROS-HE-podocitopatia	Podocitopatia	Conjuntos restritos a amostras tingidas com o corante PAS, sem rebalanceamento.
PAS-amiloidose	Amiloidose	
PAS-normal	Normal	
PAS-esclerose-pura	Esclerose pura sem crescente	
PAS-hipercelularidade	Hipercelularidade	
PAS-hipercelularidade-pura	Hipercelularidade pura sem crescente	
PAS-glomerulonefrite	Glomerulonefrite crescente	
PAS-membranosa	Nefropatia membranosa	
PAS-esclerose	Esclerose	
PAS-podocitopatia	Podocitopatia	
ROS-PAS-amiloidose	Amiloidose	
ROS-PAS-normal	Normal	

Continua na próxima página
 Conjuntos restritos a amostras tingidas com o corante PAS, gerados com rebalanceamento ROS.

Tabela A.1 – continuação da página anterior

Conjunto	Classes de interesse	Detalhes
ROS-PAS-esclerose-pura	Esclerose pura sem crescente	
ROS-PAS-hipercelularidade	Hipercelularidade	
ROS-PAS-hipercelularidade-pura	Hipercelularidade pura sem crescente	
ROS-PAS-glomerulonefrite	Glomerulonefrite crescente	
ROS-PAS-membranosa	Nefropatia membranosa	
ROS-PAS-esclerose	Esclerose	
ROS-PAS-podocitopatia	Podocitopatia	Conjuntos tratados no processo de curadoria, contendo todas as classes e corantes, sem rebalanceamento.
Curated-amiloidose	Amiloidose	
Curated-normal	Normal	
Curated-esclerose-pura	Esclerose pura sem crescente	
Curated-hipercelularidade	Hipercelularidade	
Curated-hipercelularidade-pura	Hipercelularidade pura sem crescente	
Curated-glomerulonefrite	Glomerulonefrite crescente	
Curated-membranosa	Nefropatia membranosa	
Curated-esclerose	Esclerose	
Curated-podocitopatia	Podocitopatia	Conjuntos tratados no processo de curadoria, contendo todas as classes e corantes, gerados com rebalanceamento ROS.
ROS-Curated-amiloidose	Amiloidose	
ROS-Curated-normal	Normal	
ROS-Curated-esclerose-pura	Esclerose pura sem crescente	
ROS-Curated-hipercelularidade	Hipercelularidade	
ROS-Curated-hipercelularidade-pura	Hipercelularidade pura sem crescente	
ROS-Curated-glomerulonefrite	Glomerulonefrite crescente	
ROS-Curated-membranosa	Nefropatia membranosa	
ROS-Curated-esclerose	Esclerose	
ROS-Curated-podocitopatia	Podocitopatia	
Curated-HE-amiloidose	Amiloidose	
Curated-HE-normal	Normal	

Continua na próxima página

Conjuntos tratados no processo de curadoria, restritos ao corante HE, sem rebalanceamento.

Tabela A.1 – continuação da página anterior

Conjunto	Classes de interesse	Detalhes
Curated-HE-esclerose-pura	Esclerose pura sem crescente	
Curated-HE-hipercelularidade	Hipercelularidade	
Curated-HE-hipercelularidade-pura	Hipercelularidade pura sem crescente	
Curated-HE-glomerulonefrite	Glomerulonefrite crescente	
Curated-HE-membranosa	Nefropatia membranosa	
Curated-HE-esclerose	Esclerose	
Curated-HE-podocitopatia	Podocitopatia	Conjuntos tratados no processo de curadoria, restritos ao corante HE, gerados com rebalanceamento ROS.
ROS-Curated-HE-amiloidose	Amiloidose	
ROS-Curated-HE-normal	Normal	
ROS-Curated-HE-esclerose-pura	Esclerose pura sem crescente	
ROS-Curated-HE-hipercelularidade	Hipercelularidade	
ROS-Curated-HE-hipercelularidade-pura	Hipercelularidade pura sem crescente	
ROS-Curated-HE-glomerulonefrite	Glomerulonefrite crescente	
ROS-Curated-HE-membranosa	Nefropatia membranosa	
ROS-Curated-HE-esclerose	Esclerose	
ROS-Curated-HE-podocitopatia	Podocitopatia	
Curated-PAS-amiloidose	Amiloidose	Conjuntos tratados no processo de curadoria, restritos ao corante PAS, sem rebalanceamento.
Curated-PAS-normal	Normal	
Curated-PAS-esclerose-pura	Esclerose pura sem crescente	
Curated-PAS-hipercelularidade	Hipercelularidade	
Curated-PAS-hipercelularidade-pura	Hipercelularidade pura sem crescente	
Curated-PAS-glomerulonefrite	Glomerulonefrite crescente	
Continua na próxima página		

Tabela A.1 – continuação da página anterior

Conjunto	Classes de interesse	Detalhes
Curated-PAS-membranosa	Nefropatia membranosa	
Curated-PAS-esclerose	Esclerose	
Curated-PAS-podocitopatia	Podocitopatia	
ROS-Curated-PAS-amiloidose	Amiloidose	Conjuntos tratados no processo de curadoria, restritos ao corante PAS, gerados com rebalanceamento ROS.
ROS-Curated-PAS-normal	Normal	
ROS-Curated-PAS-esclerose-pura	Esclerose pura sem crescente	
ROS-Curated-PAS-hipercelularidade	Hipercelularidade	
ROS-Curated-PAS-hipercelularidade-pura	Hipercelularidade pura sem crescente	
ROS-Curated-PAS-glomerulonefrite	Glomerulonefrite crescente	
ROS-Curated-PAS-membranosa	Nefropatia membranosa	
ROS-Curated-PAS-esclerose	Esclerose	
ROS-Curated-PAS-podocitopatia	Podocitopatia	

Apêndice B

Pesos Utilizados para o Aprendizado Sensível ao Custo

Tabela B.1: Pesos utilizados para os conjuntos *Baseline-[nome da classe]* e *ROS-[nome da classe]*.

Classe de interesse	Peso cl. negativa	Peso cl. positiva	IR
Amiloidose	0,032	0,968	$\approx 1 : 30$
Normal	0,230	0,770	$\approx 1 : 3$
Esclerose pura sem crescente	0,126	0,874	$\approx 1 : 7$
Hipercelularidade	0,268	0,732	$\approx 1 : 3$
Hipercelularidade pura	0,019	0,981	$\approx 1 : 51$
Glomerulonefrite crescente	0,094	0,906	$\approx 1 : 10$
Nefropatia membranosa	0,131	0,869	$\approx 1 : 7$
Esclerose	0,053	0,947	$\approx 1 : 18$
Podocitopatia	0,043	0,957	$\approx 1 : 22$

Tabela B.2: Pesos utilizados para os conjuntos *HE-[nome da classe]* e *ROS-HE-[nome da classe]*.

Classe de interesse	Peso cl. negativa	Peso cl. positiva	IR
Amiloidose	0,025	0,975	$\approx 1 : 39$
Normal	0,273	0,727	$\approx 1 : 3$
Esclerose pura sem crescente	0,115	0,885	$\approx 1 : 8$
Hipercelularidade	0,325	0,675	$\approx 1 : 2$
Glomerulonefrite crescente	0,080	0,920	$\approx 1 : 11$
Nefropatia membranosa	0,122	0,878	$\approx 1 : 7$
Esclerose	0,047	0,953	$\approx 1 : 20$
Podocitopatia	0,011	0,989	$\approx 1 : 88$

Tabela B.3: Pesos utilizados para os conjuntos *PAS-[nome da classe]* e *ROS-PAS-[nome da classe]*.

Classe de interesse	Peso cl. negativa	Peso cl. positiva	IR
Amiloidose	0,029	0,971	$\approx 1 : 33$
Normal	0,156	0,844	$\approx 1 : 5$
Esclerose pura sem crescente	0,136	0,864	$\approx 1 : 6$
Hipercelularidade	0,285	0,715	$\approx 1 : 3$
Hipercelularidade pura	0,047	0,953	$\approx 1 : 20$
Glomerulonefrite crescente	0,103	0,897	$\approx 1 : 9$
Nefropatia membranosa	0,106	0,894	$\approx 1 : 8$
Esclerose	0,063	0,937	$\approx 1 : 15$
Podocitopatia	0,070	0,930	$\approx 1 : 13$

Tabela B.4: Pesos utilizados para os conjuntos *Curated-[nome da classe]* e *ROS-Curated-[nome da classe]*.

Classe de interesse	Peso cl. negativa	Peso cl. positiva	IR
Amiloidose	0,029	0,971	$\approx 1 : 32$
Normal	0,233	0,767	$\approx 1 : 4$
Esclerose pura sem crescente	0,118	0,882	$\approx 1 : 7$
Hipercelularidade	0,272	0,728	$\approx 1 : 3$
Hipercelularidade pura	0,017	0,983	$\approx 1 : 55$
Glomerulonefrite crescente	0,096	0,904	$\approx 1 : 9$
Nefropatia membranosa	0,140	0,860	$\approx 1 : 6$
Esclerose	0,049	0,951	$\approx 1 : 19$
Podocitopatia	0,040	0,960	$\approx 1 : 23$

Tabela B.5: Pesos utilizados para os conjuntos *Curated-HE-[nome da classe]* e *ROS-Curated-HE-[nome da classe]*.

Classe de interesse	Peso cl. negativa	Peso cl. positiva	IR
Amiloidose	0,022	0,978	$\approx 1 : 43$
Normal	0,271	0,729	$\approx 1 : 3$
Esclerose pura sem crescente	0,106	0,894	$\approx 1 : 8$
Hipercelularidade	0,327	0,673	$\approx 1 : 2$
Glomerulonefrite crescente	0,082	0,918	$\approx 1 : 11$
Nefropatia membranosa	0,134	0,866	$\approx 1 : 6$
Esclerose	0,044	0,956	$\approx 1 : 22$
Podocitopatia	0,010	0,990	$\approx 1 : 96$

Tabela B.6: Pesos utilizados para os conjuntos *Curated-PAS-[nome da classe]* e *ROS-Curated-PAS-[nome da classe]*.

Classe de interesse	Peso cl. negativa	Peso cl. positiva	IR
Amiloidose	0,027	0,973	$\approx 1 : 35$
Normal	0,158	0,842	$\approx 1 : 5$
Esclerose pura sem crescente	0,128	0,872	$\approx 1 : 7$
Hipercelularidade	0,291	0,709	$\approx 1 : 2$
Hipercelularidade pura	0,044	0,956	$\approx 1 : 21$
Glomerulonefrite crescente	0,103	0,897	$\approx 1 : 9$
Nefropatia membranosa	0,117	0,883	$\approx 1 : 7$
Esclerose	0,059	0,941	$\approx 1 : 16$
Podocitopatia	0,066	0,934	$\approx 1 : 14$

Apêndice C

Relação Completa dos Resultados Alcançados

C.1 Resultados para os Modelos Treinados com Métodos Comuns sobre Conjuntos sem Curadoria e sem Rebalanceamento

Esta seção apresenta resultados para modelos baseados em CNN treinados com aprendizado sensível ao custo sobre conjuntos de dados sem rebalanceamento. Os pesos utilizados para cada treinamento encontram-se nas Tabelas B.1 a B.6.

C.1.1 Resultados para os Modelos Treinados com Conjuntos sem Curadoria e sem Restrição por Corante

Aqui são descritos os resultados e as observações feitas na construção dos modelos baseados .

Tabela C.1: Resultados dos experimentos realizados sobre o conjunto *Baseline-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,821($\pm 0,024$)	0,928($\pm 0,012$)	0,871($\pm 0,012$)
ResNet-18	0,772($\pm 0,031$)	0,890($\pm 0,056$)	0,825($\pm 0,016$)
ResNet-50	0,763($\pm 0,050$)	0,898($\pm 0,016$)	0,824($\pm 0,033$)

Tabela C.2: Resultados dos experimentos realizados sobre o conjunto *Baseline-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,916($\pm 0,009$)	0,967($\pm 0,006$)	0,941($\pm 0,004$)
VGG-16	0,905($\pm 0,007$)	0,965($\pm 0,009$)	0,934($\pm 0,004$)
ResNet-18	0,896($\pm 0,015$)	0,963($\pm 0,011$)	0,928($\pm 0,009$)

Tabela C.3: Resultados dos experimentos realizados sobre o conjunto *Baseline-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,355($\pm 0,010$)	0,869($\pm 0,023$)	0,504($\pm 0,013$)
VGG-16	0,37($\pm 0,015$)	0,803($\pm 0,022$)	0,502($\pm 0,018$)
EfficientNet-B0	0,342($\pm 0,007$)	0,791($\pm 0,041$)	0,477($\pm 0,012$)

Tabela C.4: Resultados dos experimentos realizados sobre o conjunto *Baseline-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,620($\pm 0,008$)	0,906($\pm 0,008$)	0,736($\pm 0,006$)
VGG-16	0,628($\pm 0,009$)	0,879($\pm 0,009$)	0,732($\pm 0,008$)
Inception-V3	0,621($\pm 0,008$)	0,843($\pm 0,014$)	0,715($\pm 0,009$)

Tabela C.5: Resultados dos experimentos realizados sobre o conjunto *Baseline-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,148($\pm 0,013$)	0,598($\pm 0,063$)	0,238($\pm 0,021$)
EfficientNet-B0	0,150($\pm 0,026$)	0,553($\pm 0,086$)	0,236($\pm 0,039$)
ResNet-18	0,149($\pm 0,029$)	0,544($\pm 0,143$)	0,233($\pm 0,049$)

Tabela C.6: Resultados dos experimentos realizados sobre o conjunto *Baseline-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,757($\pm 0,026$)	0,869($\pm 0,028$)	0,809($\pm 0,024$)
ResNet-50	0,753($\pm 0,028$)	0,866($\pm 0,013$)	0,805($\pm 0,014$)
VGG-16	0,735($\pm 0,018$)	0,872($\pm 0,017$)	0,798($\pm 0,016$)

Tabela C.7: Resultados dos experimentos realizados sobre o conjunto *Baseline-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,773($\pm 0,019$)	0,887($\pm 0,024$)	0,826($\pm 0,011$)
ResNet-18	0,753($\pm 0,032$)	0,873($\pm 0,033$)	0,808($\pm 0,028$)
VGG-16	0,754($\pm 0,027$)	0,858($\pm 0,026$)	0,802($\pm 0,021$)

Tabela C.8: Resultados dos experimentos realizados sobre o conjunto *Baseline-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,597($\pm 0,046$)	0,789($\pm 0,050$)	0,678($\pm 0,031$)
ResNet-50	0,576($\pm 0,031$)	0,817($\pm 0,027$)	0,675($\pm 0,017$)
ResNet-18	0,554($\pm 0,026$)	0,765($\pm 0,055$)	0,642($\pm 0,033$)

Tabela C.9: Resultados dos experimentos realizados sobre o conjunto *Baseline-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,701($\pm 0,028$)	0,804($\pm 0,066$)	0,748($\pm 0,043$)
ResNet-50	0,689($\pm 0,018$)	0,786($\pm 0,043$)	0,733($\pm 0,023$)
Inception-V3	0,690($\pm 0,032$)	0,784($\pm 0,045$)	0,733($\pm 0,007$)

C.1.2 Resultados para os Modelos Treinados com Conjuntos sem Curadoria e Restritos ao Corante HE

Tabela C.10: Resultados dos experimentos realizados sobre o conjunto *HE-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,871($\pm 0,076$)	0,952($\pm 0,052$)	0,909($\pm 0,057$)
ResNet-18	0,829($\pm 0,085$)	0,945($\pm 0,072$)	0,882($\pm 0,071$)
ResNet-50	0,824($\pm 0,103$)	0,903($\pm 0,051$)	0,858($\pm 0,053$)

Tabela C.11: Resultados dos experimentos realizados sobre o conjunto *HE-normal*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,940($\pm 0,013$)	0,971(0,008)	0,955($\pm 0,008$)
KCN	0,918($\pm 0,019$)	0,980($\pm 0,007$)	0,948($\pm 0,008$)
ResNet-18	0,930($\pm 0,023$)	0,968($\pm 0,016$)	0,948($\pm 0,010$)

Tabela C.12: Resultados dos experimentos realizados sobre o conjunto *HE-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,340($\pm 0,012$)	0,899($\pm 0,032$)	0,494($\pm 0,014$)
VGG-16	0,345($\pm 0,009$)	0,818($\pm 0,028$)	0,485($\pm 0,006$)
EfficientNet-B0	0,313($\pm 0,015$)	0,728($\pm 0,036$)	0,437($\pm 0,017$)

Tabela C.13: Resultados dos experimentos realizados sobre o conjunto *HE-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,655($\pm 0,008$)	0,917($\pm 0,013$)	0,764($\pm 0,006$)
VGG-16	0,652($\pm 0,010$)	0,867($\pm 0,022$)	0,744($\pm 0,011$)
EfficientNet-B0	0,629($\pm 0,011$)	0,851($\pm 0,020$)	0,724($\pm 0,011$)

Tabela C.14: Resultados dos experimentos realizados sobre o conjunto *HE-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,757($\pm 0,043$)	0,893($\pm 0,022$)	0,819($\pm 0,030$)
ResNet-18	0,767($\pm 0,025$)	0,852($\pm 0,070$)	0,806($\pm 0,038$)
KCN	0,695($\pm 0,052$)	0,916($\pm 0,045$)	0,788($\pm 0,023$)

Tabela C.15: Resultados dos experimentos realizados sobre o conjunto *HE-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,787($\pm 0,049$)	0,896($\pm 0,033$)	0,837($\pm 0,027$)
ResNet-18	0,758($\pm 0,005$)	0,888($\pm 0,038$)	0,817($\pm 0,019$)
ResNet-50	0,747($\pm 0,023$)	0,879($\pm 0,043$)	0,807($\pm 0,006$)

Tabela C.16: Resultados dos experimentos realizados sobre o conjunto *HE-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,617($\pm 0,088$)	0,819($\pm 0,093$)	0,702($\pm 0,080$)
ResNet-50	0,590($\pm 0,039$)	0,775($\pm 0,048$)	0,669($\pm 0,029$)
ResNet-18	0,605($\pm 0,016$)	0,739($\pm 0,063$)	0,665($\pm 0,029$)

Tabela C.17: Resultados dos experimentos realizados sobre o conjunto *HE-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,479($\pm 0,114$)	0,646($\pm 0,042$)	0,541($\pm 0,082$)
Inception-V3	0,377($\pm 0,135$)	0,600($\pm 0,138$)	0,452($\pm 0,121$)
ResNet-18	0,374($\pm 0,109$)	0,585($\pm 0,088$)	0,452($\pm 0,097$)

C.1.3 Resultados para os Modelos Treinados com Conjuntos sem Curadoria e Restritos ao Corante PAS

Tabela C.18: Resultados dos experimentos realizados sobre o conjunto *PAS-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,838($\pm 0,088$)	0,871($\pm 0,077$)	0,849($\pm 0,033$)
ResNet-18	0,762($\pm 0,134$)	0,882($\pm 0,087$)	0,810($\pm 0,070$)
ResNet-50	0,752($\pm 0,050$)	0,832($\pm 0,076$)	0,787($\pm 0,023$)

Tabela C.19: Resultados dos experimentos realizados sobre o conjunto *PAS-normal*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,901($\pm 0,034$)	0,950($\pm 0,023$)	0,924($\pm 0,025$)
Inception-V3	0,887($\pm 0,025$)	0,950($\pm 0,027$)	0,922($\pm 0,018$)
ResNet-18	0,887($\pm 0,026$)	0,930($\pm 0,027$)	0,907($\pm 0,012$)

Tabela C.20: Resultados dos experimentos realizados sobre o conjunto *PAS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,321($\pm 0,011$)	0,843($\pm 0,038$)	0,465($\pm 0,017$)
VGG-16	0,329($\pm 0,009$)	0,735($\pm 0,035$)	0,454($\pm 0,015$)
EfficientNet-B0	0,316($\pm 0,013$)	0,710($\pm 0,104$)	0,436($\pm 0,033$)

Tabela C.21: Resultados dos experimentos realizados sobre o conjunto *PAS-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,590($\pm 0,020$)	0,862($\pm 0,023$)	0,700($\pm 0,016$)
VGG-16	0,601($\pm 0,013$)	0,821($\pm 0,030$)	0,694($\pm 0,015$)
EfficientNet-B0	0,569($\pm 0,011$)	0,845($\pm 0,024$)	0,680($\pm 0,012$)

Tabela C.22: Resultados dos experimentos realizados sobre o conjunto *PAS-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,174($\pm 0,022$)	0,689($\pm 0,022$)	0,277($\pm 0,030$)
Inception-V3	0,168($\pm 0,029$)	0,548($\pm 0,076$)	0,258($\pm 0,041$)
ResNet-18	0,168($\pm 0,015$)	0,548($\pm 0,125$)	0,256($\pm 0,030$)

Tabela C.23: Resultados dos experimentos realizados sobre o conjunto *PAS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,751($\pm 0,030$)	0,797($\pm 0,026$)	0,773($\pm 0,023$)
ResNet-18	0,708($\pm 0,060$)	0,830($\pm 0,052$)	0,763($\pm 0,048$)
ResNet-50	0,724($\pm 0,038$)	0,799($\pm 0,080$)	0,758($\pm 0,043$)

Tabela C.24: Resultados dos experimentos realizados sobre o conjunto *PAS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,772($\pm 0,042$)	0,804($\pm 0,064$)	0,786($\pm 0,041$)
ResNet-18	0,711($\pm 0,050$)	0,820($\pm 0,096$)	0,761($\pm 0,063$)
VGG-16	0,740($\pm 0,051$)	0,788($\pm 0,073$)	0,760($\pm 0,035$)

Tabela C.25: Resultados dos experimentos realizados sobre o conjunto *PAS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,645($\pm 0,078$)	0,730($\pm 0,067$)	0,681($\pm 0,041$)
ResNet-18	0,615($\pm 0,062$)	0,735($\pm 0,093$)	0,663($\pm 0,031$)
ResNet-50	0,554($\pm 0,043$)	0,776($\pm 0,018$)	0,646($\pm 0,035$)

Tabela C.26: Resultados dos experimentos realizados sobre o conjunto *PAS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,708($\pm 0,057$)	0,762($\pm 0,077$)	0,733($\pm 0,058$)
Inception-V3	0,662($\pm 0,058$)	0,758($\pm 0,071$)	0,706($\pm 0,056$)
ResNet-50	0,571($\pm 0,025$)	0,787($\pm 0,085$)	0,661($\pm 0,043$)

C.2 Resultados dos Modelos Treinados sobre os Conjuntos Curados e sem Superamostragem

Esta seção apresenta os resultados obtidos pelos modelos baseados em CNN, treinados com os conjuntos refinados pelo processo de curadoria, sem superamostragem, utilizando aprendizado sensível ao custo.

C.2.1 Resultados para os Modelos Treinados com Conjuntos com Curadoria e sem Restrição de Corante

Tabela C.27: Resultados dos experimentos realizados sobre o conjunto *Curated-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,825($\pm 0,030$)	0,920($\pm 0,037$)	0,870($\pm 0,026$)
ResNet-50	0,756($\pm 0,094$)	0,858($\pm 0,037$)	0,800($\pm 0,049$)
ResNet-18	0,723($\pm 0,023$)	0,867($\pm 0,048$)	0,788($\pm 0,031$)

Tabela C.28: Resultados dos experimentos realizados sobre o conjunto *Curated-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,928($\pm 0,008$)	0,976($\pm 0,006$)	0,951($\pm 0,004$)
VGG-16	0,918($\pm 0,006$)	0,974($\pm 0,009$)	0,945($\pm 0,006$)
KCN	0,899($\pm 0,010$)	0,975($\pm 0,006$)	0,936($\pm 0,008$)

Tabela C.29: Resultados dos experimentos realizados sobre o conjunto *Curated-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,297($\pm 0,011$)	0,833($\pm 0,017$)	0,438($\pm 0,013$)
VGG-16	0,310($\pm 0,011$)	0,741($\pm 0,039$)	0,437($\pm 0,017$)
EfficientNet-B0	0,282($\pm 0,004$)	0,700($\pm 0,033$)	0,401($\pm 0,006$)

Tabela C.30: Resultados dos experimentos realizados sobre o conjunto *Curated-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,637($\pm 0,004$)	0,923($\pm 0,012$)	0,754($\pm 0,006$)
VGG-16	0,646($\pm 0,003$)	0,898($\pm 0,007$)	0,752($\pm 0,003$)
EfficientNet-B0	0,629($\pm 0,007$)	0,906($\pm 0,011$)	0,743($\pm 0,002$)

Tabela C.31: Resultados dos experimentos realizados sobre o conjunto *Curated-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,141($\pm 0,016$)	0,558($\pm 0,076$)	0,224($\pm 0,026$)
ResNet-18	0,140($\pm 0,017$)	0,504($\pm 0,078$)	0,220($\pm 0,027$)
EfficientNet-B0	0,134($\pm 0,010$)	0,513($\pm 0,049$)	0,213($\pm 0,015$)

Tabela C.32: Resultados dos experimentos realizados sobre o conjunto *Curated-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,820($\pm 0,033$)	0,863($\pm 0,031$)	0,840($\pm 0,007$)
ResNet-18	0,795($\pm 0,017$)	0,873($\pm 0,013$)	0,832($\pm 0,012$)
VGG-16	0,759($\pm 0,010$)	0,886($\pm 0,028$)	0,817($\pm 0,011$)

Tabela C.33: Resultados dos experimentos realizados sobre o conjunto *Curated-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,833($\pm 0,029$)	0,897($\pm 0,012$)	0,863($\pm 0,011$)
ResNet-18	0,815($\pm 0,047$)	0,883($\pm 0,024$)	0,847($\pm 0,020$)
VGG-16	0,807($\pm 0,017$)	0,887($\pm 0,019$)	0,845($\pm 0,010$)

Tabela C.34: Resultados dos experimentos realizados sobre o conjunto *Curated-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,570($\pm 0,009$)	0,800($\pm 0,039$)	0,665($\pm 0,013$)
Inception-V3	0,581($\pm 0,012$)	0,774($\pm 0,010$)	0,664($\pm 0,007$)
ResNet-50	0,522($\pm 0,025$)	0,778($\pm 0,039$)	0,624($\pm 0,018$)

Tabela C.35: Resultados dos experimentos realizados sobre o conjunto *Curated-podocitopatía*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,625($\pm 0,046$)	0,765($\pm 0,021$)	0,687($\pm 0,033$)
ResNet-50	0,638($\pm 0,035$)	0,737($\pm 0,023$)	0,683($\pm 0,023$)
ResNet-18	0,621($\pm 0,027$)	0,757($\pm 0,047$)	0,681($\pm 0,020$)

C.2.2 Resultados para os Modelos Treinados com Conjuntos com Curadoria e Restritos ao Corante HE

Tabela C.36: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,834($\pm 0,082$)	0,938($\pm 0,038$)	0,882($\pm 0,061$)
ResNet-50	0,832($\pm 0,088$)	0,903($\pm 0,029$)	0,864($\pm 0,047$)
ResNet-18	0,806($\pm 0,107$)	0,890($\pm 0,066$)	0,843($\pm (\pm 0,076)$)

Tabela C.37: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,939($\pm 0,017$)	0,982($\pm 0,005$)	0,960($\pm 0,008$)
VGG-16	0,932($\pm 0,008$)	0,985($\pm 0,006$)	0,958($\pm 0,006$)
ResNet-18	0,942($\pm 0,017$)	0,974($\pm 0,010$)	0,958($\pm 0,009$)

Tabela C.38: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,029($\pm 0,017$)	0,763($\pm 0,053$)	0,415($\pm 0,026$)
KCN	0,261($\pm 0,011$)	0,845($\pm 0,037$)	0,399($\pm 0,016$)
EfficientNet-B0	0,246($\pm 0,016$)	0,622($\pm 0,088$)	0,352($\pm 0,030$)

Tabela C.39: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-hiperplasticidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,683($\pm 0,011$)	0,924($\pm 0,010$)	0,785($\pm 0,010$)
VGG-16	0,681($\pm 0,013$)	0,891($\pm 0,016$)	0,772($\pm 0,013$)
EfficientNet-B0	0,665($\pm 0,007$)	0,881($\pm 0,018$)	0,758($\pm 0,006$)

Tabela C.40: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,807($\pm 0,029$)	0,869($\pm 0,008$)	0,837($\pm 0,015$)
ResNet-18	0,784($\pm 0,045$)	0,873($\pm 0,021$)	0,826($\pm 0,034$)
ResNet-50	0,769($\pm 0,031$)	0,858($\pm 0,053$)	0,809($\pm 0,017$)

Tabela C.41: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,829($\pm 0,055$)	0,908($\pm 0,025$)	0,865($\pm 0,025$)
ResNet-18	0,832($\pm 0,021$)	0,894($\pm 0,014$)	0,862($\pm 0,016$)
VGG-16	0,815($\pm 0,018$)	0,897($\pm 0,019$)	0,853($\pm 0,007$)

Tabela C.42: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,645($\pm 0,046$)	0,774($\pm 0,047$)	0,703($\pm 0,041$)
Inception-V3	0,631($\pm 0,025$)	0,746($\pm 0,025$)	0,683($\pm 0,029$)
ResNet-50	0,626($\pm 0,069$)	0,724($\pm 0,091$)	0,667($\pm 0,046$)

Tabela C.43: Resultados dos experimentos realizados sobre o conjunto *Curated-HE-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,414($\pm 0,053$)	0,615($\pm 0,109$)	0,492($\pm 0,062$)
Inception-V3	0,392($\pm 0,082$)	0,631($\pm 0,148$)	0,482($\pm 0,102$)
ResNet-50	0,446($\pm 0,090$)	0,554($\pm 0,138$)	0,481($\pm 0,070$)

C.2.3 Resultados para os Modelos Treinados com Conjuntos com Curadoria e Restritos ao Corante PAS

Tabela C.44: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,808($\pm 0,051$)	0,882($\pm 0,044$)	0,841($\pm 0,013$)
ResNet-18	0,844($\pm 0,122$)	0,823($\pm 0,031$)	0,830($\pm 0,062$)
ResNet-50	0,744($\pm 0,081$)	0,841($\pm 0,098$)	0,788($\pm 0,078$)

Tabela C.45: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,886($\pm 0,045$)	0,971($\pm 0,013$)	0,928($\pm 0,026$)
VGG-16	0,900($\pm 0,033$)	0,953($\pm 0,025$)	0,926($\pm 0,026$)
ResNet-18	0,885($\pm 0,038$)	0,958($\pm 0,014$)	0,920($\pm 0,024$)

Tabela C.46: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,267($\pm 0,018$)	0,731($\pm 0,040$)	0,391($\pm 0,022$)
VGG-16	0,278($\pm 0,019$)	0,627($\pm 0,059$)	0,385($\pm 0,025$)
EfficientNet-B0	0,265($\pm 0,009$)	0,648($\pm 0,021$)	0,376($\pm 0,012$)

Tabela C.47: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,609($\pm 0,007$)	0,882($\pm 0,012$)	0,721($\pm 0,007$)
VGG-16	0,617($\pm 0,022$)	0,860($\pm 0,034$)	0,719($\pm 0,023$)
EfficientNet-B0	0,586($\pm 0,017$)	0,856($\pm 0,040$)	0,696($\pm 0,024$)

Tabela C.48: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,149($\pm 0,030$)	0,562($\pm 0,170$)	0,235($\pm 0,052$)
EfficientNet-B0	0,143($\pm 0,009$)	0,610($\pm 0,031$)	0,232($\pm 0,013$)
ResNet-18	0,135($\pm 0,017$)	0,493($\pm 0,061$)	0,212($\pm 0,024$)

Tabela C.49: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,805($\pm 0,047$)	0,845($\pm 0,044$)	0,823($\pm 0,044$)
ResNet-18	0,779($\pm 0,038$)	0,832($\pm 0,033$)	0,804($\pm 0,033$)
VGG-16	0,725($\pm 0,021$)	0,874($\pm 0,038$)	0,792($\pm 0,038$)

Tabela C.50: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,851($\pm 0,015$)	0,870($\pm 0,017$)	0,860($\pm 0,013$)
Inception-V3	0,831($\pm 0,047$)	0,856($\pm 0,057$)	0,842($\pm 0,038$)
ResNet-18	0,842($\pm 0,040$)	0,842($\pm 0,027$)	0,842($\pm 0,010$)

Tabela C.51: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,534($\pm 0,067$)	0,685($\pm 0,102$)	0,598($\pm 0,071$)
Inception-V3	0,500($\pm 0,006$)	0,681($\pm 0,047$)	0,014
ResNet-50	0,454($\pm 0,030$)	0,690($\pm 0,065$)	0,546($\pm 0,028$)

Tabela C.52: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,588($\pm 0,032$)	0,727($\pm 0,049$)	0,649($\pm 0,031$)
ResNet-50	0,569($\pm 0,060$)	0,706($\pm 0,069$)	0,628($\pm 0,047$)
Inception-V3	0,548($\pm 0,049$)	0,678($\pm 0,088$)	0,605($\pm 0,061$)

C.3 Resultados dos Modelos Treinados sobre os Conjuntos sem Curadoria e com Superamostragem

Esta seção apresenta os resultados para os modelos treinados com CNNs sobre os conjuntos submetidos à superamostragem aleatória em conjunto com aprendizado sensível ao custo.

C.3.1 Resultados Obtidos sobre os Conjuntos sem Curadoria e sem Restrição por Corante

Tabela C.53: Resultados dos experimentos realizados sobre o conjunto *ROS-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,763($\pm 0,020$)	0,891($\pm 0,070$)	0,820($\pm 0,025$)
VGG-16	0,623($\pm 0,087$)	0,912($\pm 0,034$)	0,736($\pm 0,061$)
ResNet-18	0,600($\pm 0,086$)	0,904($\pm 0,040$)	0,717($\pm 0,052$)

Tabela C.54: Resultados dos experimentos realizados sobre o conjunto *ROS-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,925($\pm 0,014$)	0,962($\pm 0,004$)	0,943($\pm 0,006$)
VGG-16	0,916($\pm 0,008$)	0,968($\pm 0,011$)	0,941($\pm 0,007$)
KCN	0,893($\pm 0,015$)	0,977($\pm 0,006$)	0,933($\pm 0,006$)

Tabela C.55: Resultados dos experimentos realizados sobre o conjunto *ROS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,345($\pm 0,002$)	0,901($\pm 0,019$)	0,499($\pm 0,004$)
VGG-16	0,355($\pm 0,004$)	0,750($\pm 0,024$)	0,482($\pm 0,009$)
EfficientNet-B0	0,314($\pm 0,008$)	0,897($\pm 0,023$)	0,465($\pm 0,011$)

Tabela C.56: Resultados dos experimentos realizados sobre o conjunto *ROS-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,616($\pm 0,003$)	0,934($\pm 0,010$)	0,742($\pm 0,004$)
VGG-16	0,625($\pm 0,004$)	0,885($\pm 0,010$)	0,733($\pm 0,005$)
EfficientNet-B0	0,588($\pm 0,006$)	0,928($\pm 0,008$)	0,720($\pm 0,006$)

Tabela C.57: Resultados dos experimentos realizados sobre o conjunto *ROS-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,151($\pm 0,008$)	0,710($\pm 0,095$)	0,248($\pm 0,010$)
ResNet-18	0,127($\pm 0,014$)	0,772($\pm 0,067$)	0,217($\pm 0,021$)
Inception-V3	0,129($\pm 0,017$)	0,526($\pm 0,096$)	0,207($\pm 0,028$)

Tabela C.58: Resultados dos experimentos realizados sobre o conjunto *ROS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,779($\pm 0,023$)	0,831($\pm 0,029$)	0,804($\pm 0,014$)
VGG-16	0,737($\pm 0,017$)	0,871($\pm 0,017$)	0,798($\pm 0,004$)
KCN	0,699($\pm 0,011$)	0,911($\pm 0,017$)	0,791($\pm 0,011$)

Tabela C.59: Resultados dos experimentos realizados sobre o conjunto *ROS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,816($\pm 0,067$)	0,835($\pm 0,051$)	0,823($\pm 0,019$)
VGG-16	0,779($\pm 0,029$)	0,866($\pm 0,020$)	0,820($\pm 0,025$)
KCN	0,706($\pm 0,029$)	0,917($\pm 0,026$)	0,797($\pm 0,022$)

Tabela C.60: Resultados dos experimentos realizados sobre o conjunto *ROS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,598($\pm 0,049$)	0,728($\pm 0,069$)	0,653($\pm 0,020$)
VGG-16	0,545($\pm 0,023$)	0,780($\pm 0,039$)	0,641($\pm 0,012$)
ResNet-18	0,530($\pm 0,026$)	0,755($\pm 0,068$)	0,621($\pm 0,019$)

Tabela C.61: Resultados dos experimentos realizados sobre o conjunto *ROS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,620($\pm 0,023$)	0,737($\pm 0,025$)	0,673($\pm 0,023$)
VGG-16	0,568($\pm 0,033$)	0,810($\pm 0,037$)	0,667($\pm 0,022$)
ResNet-18	0,496($\pm 0,020$)	0,826($\pm 0,031$)	0,619($\pm 0,016$)

C.3.2 Resultados dos Experimentos com Métodos Comuns sobre os conjuntos HE sem Curadoria e com Superamostragem

Tabela C.62: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,767($\pm 0,125$)	0,903($\pm 0,143$)	0,819($\pm 0,099$)
KCN	0,614($\pm 0,095$)	0,972($\pm 0,029$)	0,749($\pm 0,070$)
VGG-16	0,650($\pm 0,223$)	0,924($\pm 0,086$)	0,741($\pm 0,193$)

Tabela C.63: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,953($\pm 0,013$)	0,975($\pm 0,012$)	0,964($\pm 0,006$)
VGG-16	0,940($\pm 0,007$)	0,977($\pm 0,006$)	0,958($\pm 0,004$)
ResNet-18	0,941($\pm 0,024$)	0,971($\pm 0,018$)	0,955($\pm 0,007$)

Tabela C.64: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,336($\pm 0,012$)	0,903($\pm 0,042$)	0,490($\pm 0,015$)
VGG-16	0,341($\pm 0,014$)	0,769($\pm 0,033$)	0,472($\pm 0,018$)
EfficientNet-B0	0,305($\pm 0,013$)	0,876($\pm 0,047$)	0,452($\pm 0,016$)

Tabela C.65: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,652($\pm 0,010$)	0,932($\pm 0,009$)	0,768($\pm 0,009$)
VGG-16	0,655($\pm 0,009$)	0,872($\pm 0,025$)	0,748($\pm 0,012$)
EfficientNet-B0	0,623($\pm 0,010$)	0,899($\pm 0,028$)	0,736($\pm 0,009$)

Tabela C.66: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,766($\pm 0,051$)	0,824($\pm 0,025$)	0,793($\pm 0,035$)
KCN	0,660($\pm 0,038$)	0,916($\pm 0,035$)	0,766
VGG-16	0,661($\pm 0,038$)	0,871($\pm 0,036$)	0,751($\pm 0,028$)

Tabela C.67: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,769($\pm 0,039$)	0,853($\pm 0,043$)	0,809($\pm 0,035$)
VGG-16	0,736($\pm 0,041$)	0,879($\pm 0,040$)	0,800($\pm 0,010$)
ResNet-18	0,719($\pm 0,024$)	0,868($\pm 0,026$)	0,786($\pm 0,022$)

Tabela C.68: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,629($\pm 0,057$)	0,721($\pm 0,091$)	0,669($\pm 0,051$)
ResNet-18	0,513($\pm 0,048$)	0,815($\pm 0,049$)	0,628($\pm 0,041$)
VGG-16	0,467($\pm 0,076$)	0,841($\pm 0,029$)	0,597($\pm 0,062$)

Tabela C.69: Resultados dos experimentos realizados sobre o conjunto *ROS-HE-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,240($\pm 0,067$)	0,646($\pm 0,069$)	0,341($\pm 0,072$)
Inception-V3	0,195($\pm 0,052$)	0,553($\pm 0,100$)	0,287($\pm 0,067$)
ResNet-18	0,162($\pm 0,025$)	0,754($\pm 0,100$)	0,266($\pm 0,039$)

C.3.3 Resultados dos Experimentos com Métodos Comuns sobre os conjuntos PAS sem Curadoria e com Superamostragem

Tabela C.70: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,649($\pm 0,200$)	0,863($\pm 0,142$)	0,715($\pm 0,132$)
Inception-V3	0,610($\pm 0,124$)	0,765($\pm 0,170$)	0,664($\pm 0,090$)
ResNet-18	0,501($\pm 0,114$)	0,833($\pm 0,074$)	0,617($\pm 0,077$)

Tabela C.71: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,908($\pm 0,030$)	0,952($\pm 0,024$)	0,929($\pm 0,016$)
VGG-16	0,901($\pm 0,025$)	0,956($\pm 0,022$)	0,927($\pm 0,022$)
ResNet-18	0,878($\pm 0,023$)	0,941($\pm 0,025$)	0,908($\pm 0,019$)

Tabela C.72: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,310($\pm 0,012$)	0,924($\pm 0,052$)	0,464($\pm 0,013$)
VGG-16	0,336($\pm 0,015$)	0,722($\pm 0,062$)	0,459($\pm 0,025$)
ResNet-50	0,298($\pm 0,020$)	0,754($\pm 0,084$)	0,426($\pm 0,032$)

Tabela C.73: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,581($\pm 0,017$)	0,931($\pm 0,017$)	0,716($\pm 0,013$)
VGG-16	0,593($\pm 0,015$)	0,853($\pm 0,051$)	0,700($\pm 0,027$)
EfficientNet-B0	0,546($\pm 0,013$)	0,908($\pm 0,028$)	0,681($\pm 0,005$)

Tabela C.74: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,170($\pm 0,007$)	0,762($\pm 0,079$)	0,278($\pm 0,014$)
Inception-V3	0,147($\pm 0,015$)	0,798($\pm 0,048$)	0,248($\pm 0,024$)
ResNet-18	0,135($\pm 0,006$)	0,903($\pm 0,049$)	0,235($\pm 0,010$)

Tabela C.75: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,740($\pm 0,036$)	0,785($\pm 0,043$)	0,761($\pm 0,026$)
VGG-16	0,655($\pm 0,042$)	0,835($\pm 0,052$)	0,733($\pm 0,031$)
KCN	0,619($\pm 0,043$)	0,897($\pm 0,053$)	0,731($\pm 0,031$)

Tabela C.76: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,759($\pm 0,073$)	0,818($\pm 0,089$)	0,783($\pm 0,039$)
Inception-V3	0,815($\pm 0,059$)	0,728($\pm 0,105$)	0,764($\pm 0,060$)
KCN	0,613($\pm 0,070$)	0,935($\pm 0,039$)	0,739($\pm 0,054$)

Tabela C.77: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,534($\pm 0,050$)	0,776($\pm 0,021$)	0,631($\pm 0,032$)
ResNet-18	0,520($\pm 0,086$)	0,762($\pm 0,089$)	0,609($\pm 0,025$)
Inception-V3	0,560($\pm 0,090$)	0,657($\pm 0,082$)	0,603($\pm 0,083$)

Tabela C.78: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,500($\pm 0,058$)	0,848($\pm 0,060$)	0,627($\pm 0,046$)
Inception-V3	0,491($\pm 0,030$)	0,799($\pm 0,052$)	0,608($\pm 0,033$)
ResNet-18	0,450($\pm 0,015$)	0,848($\pm 0,088$)	0,587($\pm 0,030$)

C.4 Resultados para os Modelos treinados com os conjuntos Curados e com Superamostragem

Esta seção apresenta os resultados obtidos pelo treinamento de modelos baseados em CNN, sobre os conjuntos refinados pelo processo de curadoria, utilizando aprendizado sensível ao custo.

C.4.1 Resultados para os Modelos Treinados com Conjuntos Curados e sem Restrição por Corante

Tabela C.79: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,766($\pm 0,024$)	0,883($\pm 0,030$)	0,820($\pm 0,018$)
Inception-V3	0,760($\pm 0,029$)	0,848($\pm 0,036$)	0,800($\pm 0,007$)
ResNet-18	0,628($\pm 0,062$)	0,853($\pm 0,059$)	0,720($\pm 0,029$)

Tabela C.80: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,932($\pm 0,011$)	0,975($\pm 0,006$)	0,953($\pm 0,005$)
VGG-16	0,918($\pm 0,007$)	0,976($\pm 0,012$)	0,946($\pm 0,009$)
ResNet-18	0,921($\pm 0,010$)	0,968($\pm 0,004$)	0,944($\pm 0,004$)

Tabela C.81: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,286($\pm 0,003$)	0,879($\pm 0,008$)	0,432($\pm 0,004$)
VGG-16	0,300($\pm 0,013$)	0,688($\pm 0,049$)	0,418($\pm 0,022$)
EfficientNet-B0	0,259($\pm 0,003$)	0,831($\pm 0,026$)	0,395($\pm 0,004$)

Tabela C.82: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,650($\pm 0,006$)	0,912($\pm 0,014$)	0,759($\pm 0,006$)
KCN	0,637($\pm 0,011$)	0,939($\pm 0,011$)	0,759($\pm 0,008$)
EfficientNet-B0	0,610($\pm 0,010$)	0,947($\pm 0,008$)	0,742($\pm 0,006$)

Tabela C.83: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,130($\pm 0,016$)	0,697($\pm 0,078$)	0,218($\pm 0,021$)
ResNet-18	0,110($\pm 0,008$)	0,696($\pm 0,036$)	0,190($\pm 0,012$)
Inception-V3	0,109($\pm 0,017$)	0,428($\pm 0,082$)	0,173($\pm 0,029$)

Tabela C.84: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,806($\pm 0,023$)	0,866($\pm 0,011$)	0,835($\pm 0,013$)
VGG-16	0,780($\pm 0,018$)	0,880($\pm 0,015$)	0,827($\pm 0,006$)
KCN	0,728($\pm 0,028$)	0,926($\pm 0,022$)	0,815($\pm 0,016$)

Tabela C.85: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,847($\pm 0,021$)	0,886($\pm 0,020$)	0,866($\pm 0,018$)
Inception-V3	0,841($\pm 0,022$)	0,875($\pm 0,025$)	0,858($\pm 0,014$)
KCN	0,780($\pm 0,026$)	0,909($\pm 0,020$)	0,839($\pm 0,015$)

Tabela C.86: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,544($\pm 0,037$)	0,759($\pm 0,052$)	0,633($\pm 0,032$)
Inception-V3	0,576($\pm 0,034$)	0,672($\pm 0,033$)	0,620($\pm 0,020$)
KCN	0,452($\pm 0,021$)	0,874($\pm 0,017$)	0,595($\pm 0,017$)

Tabela C.87: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,547($\pm 0,088$)	0,773($\pm 0,069$)	0,634($\pm 0,048$)
ResNet-18	0,486($\pm 0,017$)	0,784($\pm 0,043$)	0,600($\pm 0,024$)
VGG-16	0,471($\pm 0,123$)	0,810($\pm 0,068$)	0,583($\pm 0,103$)

C.4.2 Resultados para os Modelos treinados com os conjuntos Curados Restritos ao Corante HE e com Superamostragem

Tabela C.88: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,730($\pm 0,087$)	0,848($\pm 0,111$)	0,783($\pm 0,091$)
VGG-16	0,673($\pm 0,196$)	0,896($\pm 0,073$)	0,759($\pm 0,168$)
ResNet-18	0,602($\pm 0,084$)	0,903($\pm 0,045$)	0,719($\pm 0,060$)

Tabela C.89: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-normal*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,947($\pm 0,020$)	0,980($\pm 0,009$)	0,963($\pm 0,009$)
VGG-16	0,936($\pm 0,009$)	0,988($\pm 0,009$)	0,961($\pm 0,002$)
KCN	0,934($\pm 0,011$)	0,987($\pm 0,007$)	0,959($\pm 0,005$)

Tabela C.90: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,258($\pm 0,007$)	0,896($\pm 0,046$)	0,400($\pm 0,010$)
EfficientNet-B0	0,239($\pm 0,005$)	0,867($\pm 0,025$)	0,374($\pm 0,006$)
VGG-16	0,254($\pm 0,019$)	0,649($\pm 0,053$)	0,365($\pm 0,028$)

Tabela C.91: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-hiperplasia*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,675($\pm 0,013$)	0,945($\pm 0,023$)	0,788($\pm 0,015$)
VGG-16	0,681($\pm 0,009$)	0,897($\pm 0,014$)	0,774($\pm 0,008$)
EfficientNet-B0	0,656($\pm 0,013$)	0,918($\pm 0,017$)	0,765($\pm 0,007$)

Tabela C.92: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,819($\pm 0,051$)	0,843($\pm 0,036$)	0,830($\pm 0,037$)
KCN	0,717($\pm 0,045$)	0,935($\pm 0,037$)	0,810($\pm 0,022$)
VGG-16	0,732($\pm 0,041$)	0,881($\pm 0,031$)	0,799($\pm 0,023$)

Tabela C.93: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,854($\pm 0,037$)	0,871($\pm 0,034$)	0,861($\pm 0,010$)
VGG-16	0,818($\pm 0,042$)	0,902($\pm 0,023$)	0,857($\pm 0,022$)
KCN	0,763($\pm 0,029$)	0,920($\pm 0,018$)	0,834($\pm 0,020$)

Tabela C.94: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,627($\pm 0,074$)	0,702($\pm 0,023$)	0,661($\pm 0,043$)
ResNet-18	0,581($\pm 0,048$)	0,713($\pm 0,060$)	0,639($\pm 0,039$)
VGG-16	0,545($\pm 0,071$)	0,763($\pm 0,074$)	0,635($\pm 0,067$)

Tabela C.95: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-HE-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,288($\pm 0,053$)	0,615($\pm 0,077$)	0,390($\pm 0,056$)
VGG-16	0,249($\pm 0,054$)	0,631($\pm 0,175$)	0,356($\pm 0,081$)
ResNet-18	0,162($\pm 0,020$)	0,800($\pm 0,117$)	0,270($\pm 0,033$)

C.4.3 Resultados para os Modelos treinados com os conjuntos Curados Restritos ao Corante PAS e com Superamostragem

Tabela C.96: Resultados dos experimentos realizados sobre o conjunto *ROS-PAS-Curated-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,654($\pm 0,125$)	0,891($\pm 0,090$)	0,748($\pm 0,095$)
Inception-V3	0,5430,059	0,793($\pm 0,098$)	0,641($\pm 0,054$)
ResNet-18	0,458($\pm 0,114$)	0,862($\pm 0,056$)	0,591($\pm 0,102$)

Tabela C.97: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,907($\pm 0,040$)	0,964($\pm 0,007$)	0,934($\pm 0,024$)
VGG-16	0,906($\pm 0,028$)	0,964($\pm 0,015$)	0,934($\pm 0,021$)
ResNet-18	0,893($\pm 0,060$)	0,955($\pm 0,031$)	0,922($\pm 0,026$)

Tabela C.98: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,244($\pm 0,009$)	0,902($\pm 0,055$)	0,384($\pm 0,010$)
VGG-16	0,277($\pm 0,013$)	0,612($\pm 0,050$)	0,381($\pm 0,017$)
EfficientNet-B0	0,223($\pm 0,010$)	0,875($\pm 0,065$)	0,355($\pm 0,018$)

Tabela C.99: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-hiper celularidade*.

Arquitetura	Precisão	Recall	F1-Score
KCN	0,615($\pm 0,009$)	0,914($\pm 0,035$)	0,735($\pm 0,013$)
VGG-16	0,620($\pm 0,016$)	0,865($\pm 0,031$)	0,722($\pm 0,021$)
EfficientNet-B0	0,580($\pm 0,013$)	0,880($\pm 0,032$)	0,699($\pm 0,016$)

Tabela C.100: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-hiper celularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,149($\pm 0,016$)	0,665($\pm 0,076$)	0,244($\pm 0,026$)
Inception-V3	0,140($\pm 0,007$)	0,750($\pm 0,068$)	0,236($\pm 0,011$)
ResNet-18	0,121($\pm 0,013$)	0,744($\pm 0,034$)	0,208($\pm 0,019$)

Tabela C.101: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,708($\pm 0,049$)	0,887($\pm 0,030$)	0,786($\pm 0,033$)
Inception-V3	0,757($\pm 0,041$)	0,805($\pm 0,058$)	0,778($\pm 0,017$)
ResNet-18	0,706($\pm 0,067$)	0,839($\pm 0,044$)	0,765($\pm 0,034$)

Tabela C.102: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,833($\pm 0,021$)	0,882($\pm 0,040$)	0,857($\pm 0,027$)
Inception-V3	0,846($\pm 0,049$)	0,835($\pm 0,033$)	0,840($\pm 0,030$)
ResNet-18	0,769($\pm 0,037$)	0,898($\pm 0,045$)	0,828($\pm 0,030$)

Tabela C.103: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,443($\pm 0,016$)	0,772($\pm 0,031$)	0,562($\pm 0,014$)
Inception-V3	0,475($\pm 0,041$)	0,653($\pm 0,038$)	0,550($\pm 0,039$)
ResNet-18	0,394($\pm 0,050$)	0,749($\pm 0,059$)	0,516($\pm 0,055$)

Tabela C.104: Resultados dos experimentos realizados sobre o conjunto *ROS-Curated-PAS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
VGG-16	0,432($\pm 0,030$)	0,816($\pm 0,072$)	0,565($\pm 0,038$)
Inception-V3	0,430($\pm 0,071$)	0,767($\pm 0,062$)	0,550($\pm 0,072$)
ResNet-18	0,405($\pm 0,034$)	0,800($\pm 0,062$)	0,537($\pm 0,038$)

C.5 Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos sem Curadoria

Esta seção apresenta os resultados obtidos pelos modelos treinados com FSL sobre os conjuntos não curados. Salienta-se o fato de que todos os experimentos com FSL foram realizados sobre conjuntos sem superamostragem e os modelos foram construídos com a aplicação de aprendizados sensível ao custo.

C.5.1 Resultados para os Modelos Treinados com os Conjuntos sem Curadoria e sem Restrição por Corante

Tabela C.105: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,774($\pm 0,064$)	0,923($\pm 0,054$)	0,840($\pm 0,036$)
Inception-V3	0,771($\pm 0,082$)	0,923($\pm 0,017$)	0,837($\pm 0,046$)
ResNet-50	0,727($\pm 0,132$)	0,939($\pm 0,027$)	0,813($\pm 0,084$)

Tabela C.106: Resultados dos experimentos com FLS realizados sobre o conjunto *Baseline-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,860($\pm 0,060$)	0,968($\pm 0,012$)	0,909($\pm 0,031$)
ResNet-50	0,824($\pm 0,053$)	0,974($\pm 0,015$)	0,892($\pm 0,027$)
EfficientNet-B0	0,807($\pm 0,041$)	0,984($\pm 0,013$)	0,886($\pm 0,029$)

Tabela C.107: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,310($\pm 0,009$)	0,909($\pm 0,023$)	0,462($\pm 0,009$)
Inception-V3	0,302($\pm 0,014$)	0,893($\pm 0,057$)	0,451($\pm 0,016$)
EfficientNet-B4	0,293($\pm 0,022$)	0,908($\pm 0,029$)	0,442($\pm 0,022$)

Tabela C.108: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,591($\pm 0,024$)	0,923($\pm 0,012$)	0,720($\pm 0,015$)
Inception-V3	0,599($\pm 0,025$)	0,891($\pm 0,042$)	0,715($\pm 0,012$)
ResNet-50	0,576($\pm 0,021$)	0,934($\pm 0,020$)	0,712($\pm 0,014$)

Tabela C.109: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,136($\pm 0,017$)	0,853($\pm 0,058$)	0,233($\pm 0,024$)
ResNet-18	0,120($\pm 0,017$)	0,875($\pm 0,034$)	0,211($\pm 0,027$)
Inception-V3	0,102($\pm 0,018$)	0,492($\pm 0,196$)	0,167($\pm 0,036$)

Tabela C.110: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,646($\pm 0,096$)	0,919($\pm 0,015$)	0,755($\pm 0,067$)
Inception-V3	0,657($\pm 0,076$)	0,893($\pm 0,028$)	0,754($\pm 0,043$)
EfficientNet-B0	0,563($\pm 0,028$)	0,941($\pm 0,018$)	0,704($\pm 0,020$)

Tabela C.111: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,675($\pm 0,092$)	0,876($\pm 0,043$)	0,757($\pm 0,044$)
ResNet-50	0,561($\pm 0,093$)	0,939($\pm 0,031$)	0,697($\pm 0,063$)
EfficientNet-B0	0,537($\pm 0,079$)	0,929($\pm 0,029$)	0,676($\pm 0,056$)

Tabela C.112: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,555($\pm 0,046$)	0,784($\pm 0,044$)	0,649($\pm 0,035$)
ResNet-50	0,424($\pm 0,059$)	0,867($\pm 0,015$)	0,567($\pm 0,053$)
EfficientNet-B0	0,384($\pm 0,047$)	0,904($\pm 0,032$)	0,538($\pm 0,045$)

Tabela C.113: Resultados dos experimentos com FSL realizados sobre o conjunto *Baseline-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,507($\pm 0,057$)	0,818($\pm 0,046$)	0,624($\pm 0,047$)
ResNet-50	0,469($\pm 0,064$)	0,877($\pm 0,029$)	0,609($\pm 0,052$)
EfficientNet-B0	0,356($\pm 0,078$)	0,911($\pm 0,020$)	0,508($\pm 0,085$)

C.5.2 Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos sem Curadoria Restritos ao Corante HE

Tabela C.114: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,904($\pm 0,040$)	0,952($\pm 0,052$)	0,925($\pm 0,031$)
EfficientNet-B0	0,840($\pm 0,061$)	0,952($\pm 0,031$)	0,891($\pm 0,037$)
Inception-V3	0,882($\pm 0,109$)	0,910($\pm 0,079$)	0,891($\pm 0,072$)

Tabela C.115: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-normal*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,926($\pm 0,006$)	0,978($\pm 0,011$)	0,952($\pm 0,005$)
Inception-V3	0,931($\pm 0,012$)	0,973($\pm 0,007$)	0,952($\pm 0,006$)
ResNet-18	0,917($\pm 0,014$)	0,980($\pm 0,007$)	0,947($\pm 0,009$)

Tabela C.116: Resultados dos experimentos com FSL realizados sobre o conjunto *FSL-HE-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,320($\pm 0,013$)	0,803($\pm 0,028$)	0,458($\pm 0,017$)
EfficientNet-B0	0,308($\pm 0,024$)	0,884($\pm 0,068$)	0,456($\pm 0,025$)
EfficientNet-B4	0,299($\pm 0,021$)	0,896($\pm 0,024$)	0,448($\pm 0,021$)

Tabela C.117: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-hiperplasticidade*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,618($\pm 0,021$)	0,925($\pm 0,064$)	0,740($\pm 0,009$)
EfficientNet-B0	0,632($\pm 0,021$)	0,880($\pm 0,014$)	0,736($\pm 0,015$)
Inception-V3	0,641($\pm 0,029$)	0,866($\pm 0,057$)	0,735($\pm 0,008$)

Tabela C.118: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,719($\pm 0,037$)	0,906($\pm 0,037$)	0,801($\pm 0,025$)
ResNet-50	0,713($\pm 0,042$)	0,904($\pm 0,044$)	0,796($\pm 0,027$)
Inception-V3	0,764($\pm 0,051$)	0,835($\pm 0,061$)	0,795($\pm 0,016$)

Tabela C.119: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,746($\pm 0,075$)	0,923($\pm 0,041$)	0,822($\pm 0,045$)
Inception-V3	0,755($\pm 0,057$)	0,879($\pm 0,029$)	0,812($\pm 0,042$)
EfficientNet-B0	0,688($\pm 0,083$)	0,916($\pm 0,046$)	0,781($\pm 0,041$)

Tabela C.120: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,722($\pm 0,093$)	0,750($\pm 0,088$)	0,732($\pm 0,074$)
ResNet-50	0,665($\pm 0,061$)	0,786($\pm 0,049$)	0,720($\pm 0,055$)
ResNet-18	0,622($\pm 0,039$)	0,811($\pm 0,113$)	0,700($\pm 0,049$)

Tabela C.121: Resultados dos experimentos com FSL realizados sobre o conjunto *HE-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,607($\pm 0,125$)	0,492($\pm 0,103$)	0,543($\pm 0,111$)
EfficientNet-B0	0,558($\pm 0,078$)	0,508($\pm 0,088$)	0,527($\pm 0,069$)
Inception-V3	0,604($\pm 0,149$)	0,446($\pm 0,034$)	0,504($\pm 0,054$)

C.5.3 Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Sem Curadoria Restritos ao Corante PAS

Tabela C.122: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,842($\pm 0,067$)	0,931($\pm 0,056$)	0,881($\pm 0,027$)
Inception-V3	0,898($\pm 0,077$)	0,814($\pm 0,086$)	0,850($\pm 0,056$)
ResNet-18	0,854($\pm 0,141$)	0,843($\pm 0,092$)	0,842($\pm 0,087$)

Tabela C.123: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-normal*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,895($\pm 0,030$)	0,967($\pm 0,014$)	0,929($\pm 0,022$)
Inception-V3	0,919($\pm 0,034$)	0,933($\pm 0,022$)	0,926($\pm 0,025$)
ResNet-18	0,878($\pm 0,047$)	0,965($\pm 0,034$)	0,919($\pm 0,024$)

Tabela C.124: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,324($\pm 0,014$)	0,754($\pm 0,048$)	0,453($\pm 0,021$)
ResNet-50	0,299($\pm 0,035$)	0,881($\pm 0,048$)	0,445($\pm 0,036$)
EfficientNet-B4	0,309($\pm 0,015$)	0,778($\pm 0,055$)	0,441($\pm 0,015$)

Tabela C.125: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-hiperplasia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,571($\pm 0,017$)	0,895($\pm 0,043$)	0,697($\pm 0,021$)
ResNet-18	0,568($\pm 0,027$)	0,870($\pm 0,034$)	0,687($\pm 0,026$)
EfficientNet-B4	0,572($\pm 0,011$)	0,855($\pm 0,022$)	0,686($\pm 0,002$)

Tabela C.126: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-hiperplasia-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,166($\pm 0,024$)	0,604($\pm 0,161$)	0,259($\pm 0,042$)
EfficientNet-B4	0,156($\pm 0,020$)	0,750($\pm 0,119$)	0,257($\pm 0,031$)
ResNet-18	0,153($\pm 0,019$)	0,561($\pm 0,203$)	0,238($\pm 0,043$)

Tabela C.127: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,757($\pm 0,042$)	0,827($\pm 0,033$)	0,790($\pm 0,030$)
EfficientNet-B0	0,727($\pm 0,056$)	0,866($\pm 0,075$)	0,787($\pm 0,033$)
Inception-V3	0,780($\pm 0,072$)	0,780($\pm 0,045$)	0,776($\pm 0,020$)

Tabela C.128: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,806($\pm 0,031$)	0,836($\pm 0,048$)	0,819($\pm 0,014$)
ResNet-18	0,773($\pm 0,074$)	0,853($\pm 0,084$)	0,806($\pm 0,035$)
Inception-V3	0,833($\pm 0,037$)	0,777($\pm 0,093$)	0,801($\pm 0,050$)

Tabela C.129: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,689($\pm 0,118$)	0,808($\pm 0,066$)	0,739($\pm 0,079$)
Inception-V3	0,695($\pm 0,058$)	0,744($\pm 0,093$)	0,713($\pm 0,035$)
ResNet-18	0,629($\pm 0,074$)	0,803($\pm 0,076$)	0,699($\pm 0,020$)

Tabela C.130: Resultados dos experimentos com FSL realizados sobre o conjunto *PAS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,721($\pm 0,040$)	0,799($\pm 0,081$)	0,758($\pm 0,058$)
Inception-V3	0,677($\pm 0,073$)	0,717($\pm 0,085$)	0,693($\pm 0,053$)
ResNet-50	0,636($\pm 0,126$)	0,770($\pm 0,085$)	0,690($\pm 0,089$)

C.6 Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Curados

Esta seção apresenta os resultados obtidos pelos modelos baseados em redes prototípicas, treinados com conjuntos curados.

C.6.1 Resultados para os Modelos Treinados com Conjuntos Curados e sem Restrição por Corante

Tabela C.131: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,786($\pm 0,108$)	0,866($\pm 0,038$)	0,818($\pm 0,052$)
ResNet-50	0,616($\pm 0,083$)	0,912($\pm 0,023$)	0,732($\pm 0,054$)
ResNet-18	0,610($\pm 0,152$)	0,930($\pm 0,031$)	0,726($\pm 0,102$)

Tabela C.132: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-normal*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,822($\pm 0,038$)	0,979($\pm 0,016$)	0,893($\pm 0,020$)
EfficientNet-B0	0,818($\pm 0,031$)	0,981($\pm 0,006$)	0,892($\pm 0,019$)
Inception-V3	0,800($\pm 0,035$)	0,985($\pm 0,010$)	0,882($\pm 0,019$)

Tabela C.133: Resultados dos experimentos realizados sobre o conjunto *Curated-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,253($\pm 0,016$)	0,908($\pm 0,022$)	0,395($\pm 0,018$)
Inception-V3	0,242($\pm 0,020$)	0,890($\pm 0,033$)	0,380($\pm 0,024$)
EfficientNet-B4	0,241($\pm 0,007$)	0,881($\pm 0,005$)	0,378($\pm 0,009$)

Tabela C.134: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,626($\pm 0,017$)	0,942($\pm 0,005$)	0,752($\pm 0,013$)
ResNet-18	0,601($\pm 0,011$)	0,934($\pm 0,007$)	0,731($\pm 0,007$)
Inception-V3	0,593($\pm 0,036$)	0,938($\pm 0,048$)	0,726($\pm 0,023$)

Tabela C.135: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,126($\pm 0,010$)	0,723($\pm 0,052$)	0,215($\pm 0,015$)
EfficientNet-B0	0,121($\pm 0,023$)	0,804($\pm 0,098$)	0,209($\pm 0,034$)
ResNet-50	0,087($\pm 0,012$)	0,920($\pm 0,066$)	0,158($\pm 0,019$)

Tabela C.136: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,662($\pm 0,085$)	0,903($\pm 0,030$)	0,760($\pm 0,045$)
ResNet-50	0,641($\pm 0,035$)	0,927($\pm 0,022$)	0,757($\pm 0,022$)
EfficientNet-B0	0,619($\pm 0,025$)	0,936($\pm 0,014$)	0,745($\pm 0,018$)

Tabela C.137: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,604($\pm 0,090$)	0,938($\pm 0,036$)	0,730($\pm 0,058$)
ResNet-18	0,607($\pm 0,040$)	0,910($\pm 0,026$)	0,727($\pm 0,029$)
Inception-V3	0,597($\pm 0,101$)	0,935($\pm 0,029$)	0,724($\pm 0,071$)

Tabela C.138: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,516($\pm 0,087$)	0,810($\pm 0,036$)	0,625($\pm 0,066$)
ResNet-50	0,361($\pm 0,054$)	0,871($\pm 0,029$)	0,507($\pm 0,051$)
Inception-V3	0,324($\pm 0,044$)	0,919($\pm 0,013$)	0,478($\pm 0,047$)

Tabela C.139: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,371($\pm 0,056$)	0,858($\pm 0,037$)	0,515($\pm 0,052$)
EfficientNet-B0	0,335($\pm 0,061$)	0,873($\pm 0,061$)	0,480($\pm 0,058$)
ResNet-18	0,324($\pm 0,086$)	0,905($\pm 0,058$)	0,470($\pm 0,091$)

C.6.2 Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Curados Restritos ao Corante HE

Tabela C.140: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,862($\pm 0,049$)	0,890($\pm 0,057$)	0,874($\pm 0,039$)
Inception-V3	0,886($\pm 0,073$)	0,862($\pm 0,073$)	0,871($\pm 0,045$)
ResNet-18	0,831($\pm 0,029$)	0,910($\pm 0,058$)	0,868($\pm 0,033$)

Tabela C.141: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,930($\pm 0,021$)	0,987($\pm 0,006$)	0,957($\pm 0,010$)
EfficientNet-B0	0,915($\pm 0,022$)	0,985($\pm 0,006$)	0,949($\pm 0,012$)
ResNet-50	0,924($\pm 0,015$)	0,973($\pm 0,013$)	0,948($\pm 0,007$)

Tabela C.142: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,249($\pm 0,009$)	0,811($\pm 0,059$)	0,381($\pm 0,015$)
Inception-V3	0,241($\pm 0,039$)	0,819($\pm 0,079$)	0,370($\pm 0,044$)
EfficientNet-B4	0,235($\pm 0,008$)	0,856($\pm 0,046$)	0,369($\pm 0,011$)

Tabela C.143: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,675($\pm 0,015$)	0,923($\pm 0,015$)	0,780($\pm 0,011$)
Inception-V3	0,668($\pm 0,012$)	0,898($\pm 0,050$)	0,766($\pm 0,022$)
ResNet-18	0,661($\pm 0,028$)	0,910($\pm 0,056$)	0,764($\pm 0,023$)

Tabela C.144: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,823($\pm 0,045$)	0,846($\pm 0,038$)	0,833($\pm 0,015$)
ResNet-18	0,760($\pm 0,052$)	0,891($\pm 0,034$)	0,819($\pm 0,032$)
ResNet-50	0,745($\pm 0,021$)	0,887($\pm 0,047$)	0,809($\pm 0,017$)

Tabela C.145: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,860($\pm 0,075$)	0,900($\pm 0,021$)	0,877($\pm 0,032$)
ResNet-50	0,829($\pm 0,046$)	0,924($\pm 0,012$)	0,873($\pm 0,025$)
ResNet-18	0,728($\pm 0,069$)	0,933($\pm 0,018$)	0,816($\pm 0,045$)

Tabela C.146: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,712($\pm 0,047$)	0,731($\pm 0,054$)	0,719($\pm 0,025$)
ResNet-50	0,656($\pm 0,025$)	0,774($\pm 0,051$)	0,710($\pm 0,035$)
ResNet-18	0,680($\pm 0,115$)	0,742($\pm 0,048$)	0,704($\pm 0,073$)

Tabela C.147: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-HE-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,508($\pm 0,094$)	0,492($\pm 0,160$)	0,497($\pm 0,123$)
EfficientNet-B0	0,382($\pm 0,073$)	0,554($\pm 0,064$)	0,450($\pm 0,066$)
ResNet-50	0,383($\pm 0,132$)	0,415($\pm 0,169$)	0,383($\pm 0,128$)

C.6.3 Resultados para os Modelos Construídos com Few-shot Learning Aplicado aos Conjuntos Curados Restritos ao Corante PAS

Tabela C.148: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-amiloidose*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-18	0,866($\pm 0,070$)	0,822($\pm 0,078$)	0,842($\pm 0,068$)
Inception-V3	0,904($\pm 0,096$)	0,774($\pm 0,047$)	0,832($\pm 0,056$)
ResNet-50	0,801($\pm 0,159$)	0,861($\pm 0,067$)	0,825($\pm 0,115$)

Tabela C.149: Resultados dos experimentos realizados sobre o conjunto *Curated-PAS-normal*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,903($\pm 0,037$)	0,976($\pm 0,018$)	0,937($\pm 0,019$)
EfficientNet-B0	0,892($\pm 0,039$)	0,976($\pm 0,011$)	0,932($\pm 0,024$)
ResNet-18	0,896($\pm 0,036$)	0,971($\pm 0,013$)	0,932($\pm 0,023$)

Tabela C.150: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-PAS-esclerose-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,270($\pm 0,035$)	0,697($\pm 0,045$)	0,387($\pm 0,032$)
ResNet-50	0,250($\pm 0,024$)	0,839($\pm 0,050$)	0,385($\pm 0,026$)
EfficientNet-B4	0,275($\pm 0,024$)	0,646($\pm 0,057$)	0,385($\pm 0,022$)

Tabela C.151: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-PAS-hipercelularidade*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,593($\pm 0,024$)	0,910($\pm 0,015$)	0,718($\pm 0,018$)
ResNet-18	0,575($\pm 0,007$)	0,915($\pm 0,030$)	0,706($\pm 0,012$)
Inception-V3	0,582($\pm 0,019$)	0,899($\pm 0,034$)	0,706($\pm 0,011$)

Tabela C.152: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-PAS-hipercelularidade-pura*.

Arquitetura	Precisão	Recall	F1-Score
EfficientNet-B0	0,142($\pm 0,022$)	0,844($\pm 0,114$)	0,241($\pm 0,027$)
ResNet-18	0,125($\pm 0,014$)	0,817($\pm 0,072$)	0,217($\pm 0,019$)
Inception-V3	0,113($\pm 0,015$)	0,866($\pm 0,042$)	0,200($\pm 0,023$)

Tabela C.153: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-PAS-glomerulonefrite*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,843($\pm 0,043$)	0,829($\pm 0,026$)	0,836($\pm 0,034$)
ResNet-18	0,813($\pm 0,060$)	0,810($\pm 0,038$)	0,811($\pm 0,041$)
Inception-V3	0,810($\pm 0,067$)	0,808($\pm 0,052$)	0,807($\pm 0,038$)

Tabela C.154: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-PAS-membranosa*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,893($\pm 0,053$)	0,851($\pm 0,039$)	0,871($\pm 0,036$)
ResNet-18	0,847($\pm 0,061$)	0,889($\pm 0,019$)	0,866($\pm 0,033$)
Inception-V3	0,869($\pm 0,063$)	0,863($\pm 0,020$)	0,865($\pm 0,030$)

Tabela C.155: Resultados dos experimentos com FSL realizados sobre o conjunto *Curated-PAS-esclerose*.

Arquitetura	Precisão	Recall	F1-Score
Inception-V3	0,525($\pm 0,068$)	0,676($\pm 0,080$)	0,584($\pm 0,021$)
ResNet-18	0,504($\pm 0,073$)	0,694($\pm 0,126$)	0,575($\pm 0,063$)
ResNet-50	0,501($\pm 0,024$)	0,662($\pm 0,035$)	0,570($\pm 0,021$)

Tabela C.156: Resultados dos experimentos FSL realizados sobre o conjunto *Curated-PAS-podocitopatia*.

Arquitetura	Precisão	Recall	F1-Score
ResNet-50	0,614($\pm 0,059$)	0,686($\pm 0,108$)	0,614($\pm 0,077$)
Inception-V3	0,597($\pm 0,105$)	0,608($\pm 0,067$)	0,597($\pm 0,074$)
ResNet-18	0,595($\pm 0,062$)	0,739($\pm 0,121$)	0,595($\pm 0,050$)

Apêndice D

Gráficos dos Modelos de Maior Desempenho

D.1 Amiloidose

D.1.1 Curvas da Função de Custo

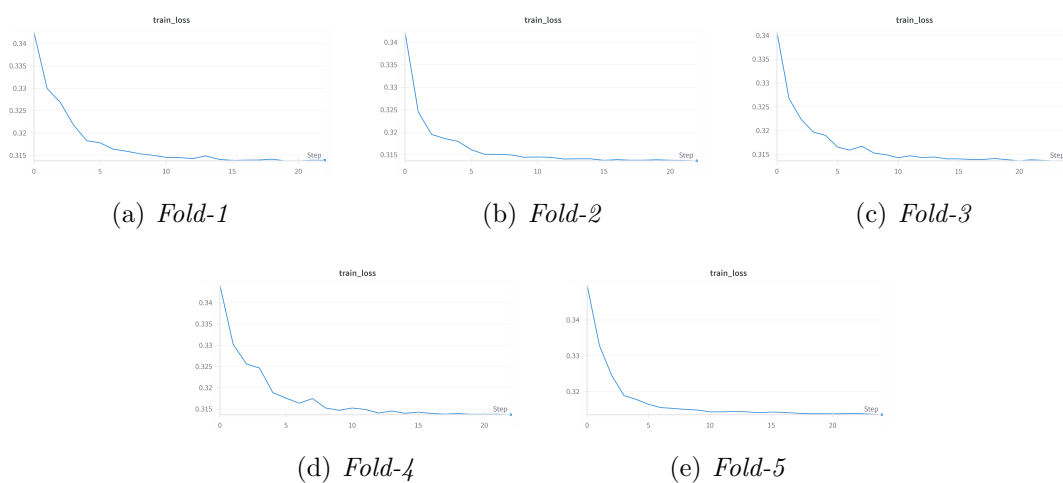


Figura D.1: Curvas da função de custo para *amiloidose*.

D.1.2 Curvas precision-recall

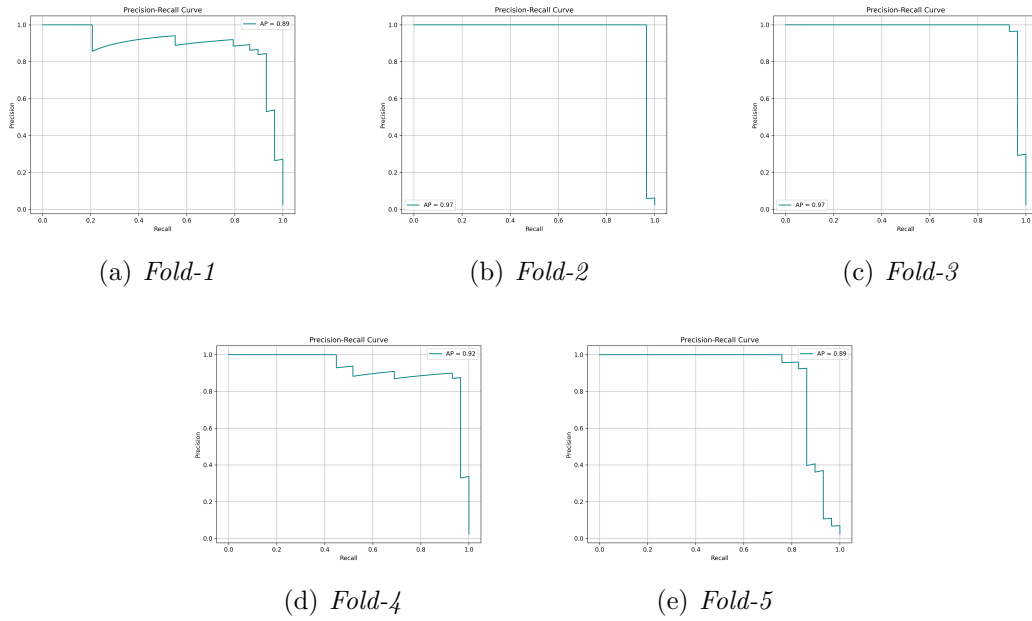


Figura D.2: Curvas *precision-recall* para *amiloidose*.

D.1.3 Matrizes de Confusão

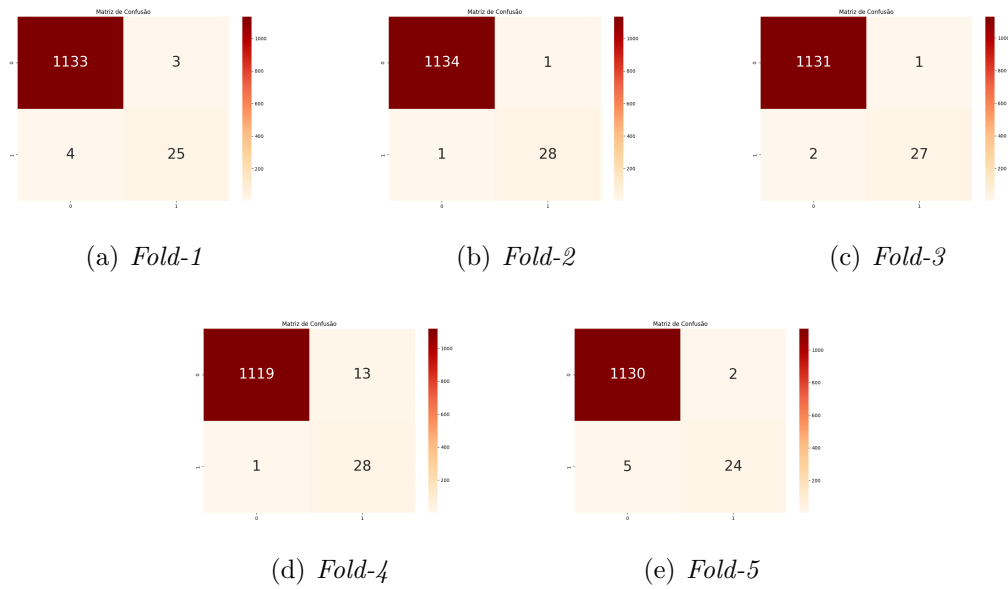


Figura D.3: Matrizes de confusão para *amiloidose*.

D.2 Normal

D.2.1 Curvas da Função de Custo

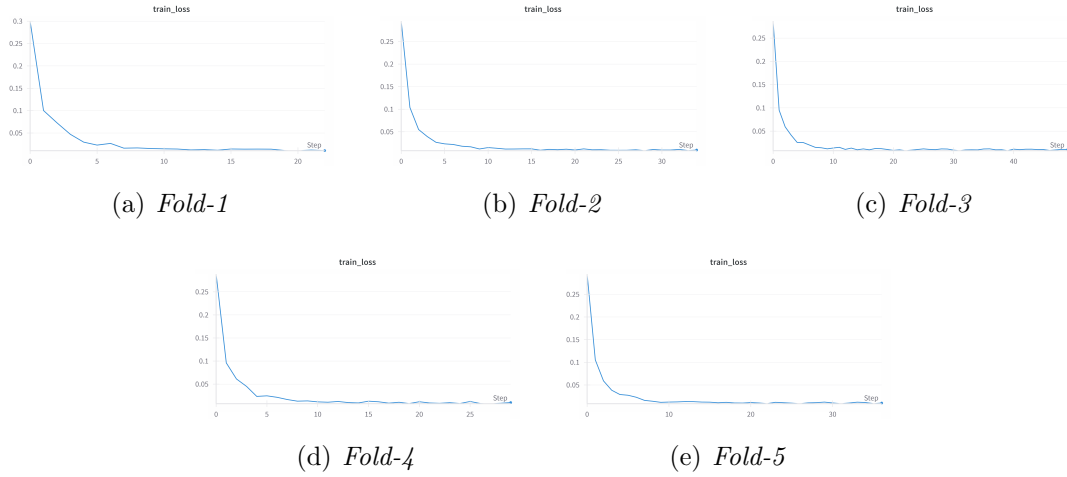


Figura D.4: Curvas da função de custo para *normal*.

D.2.2 Curvas precision-recall

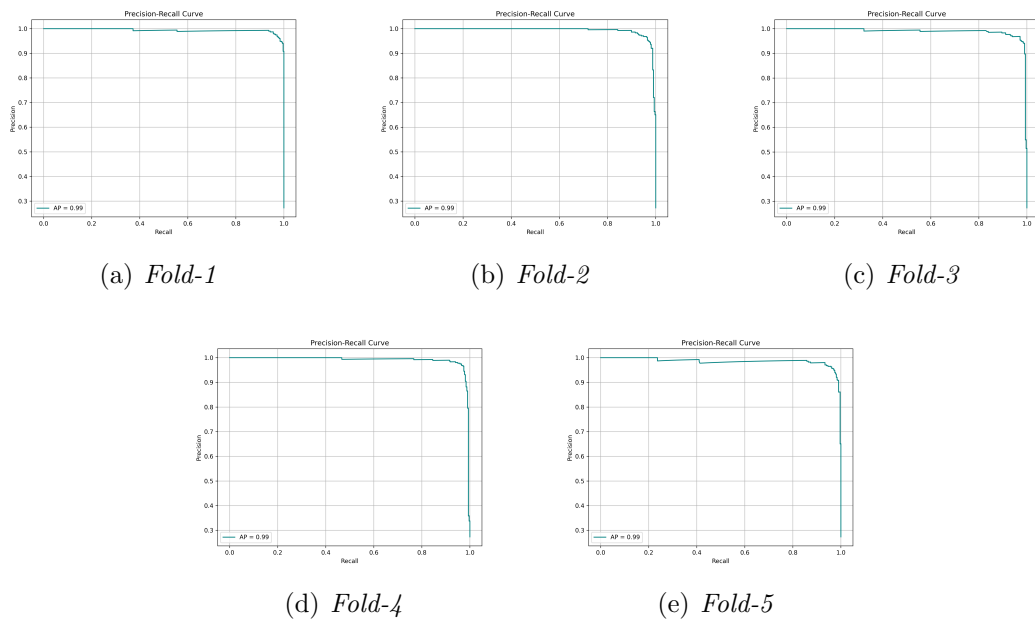


Figura D.5: Curvas *precision-recall* para *normal*.

D.2.3 Matrizes de Confusão

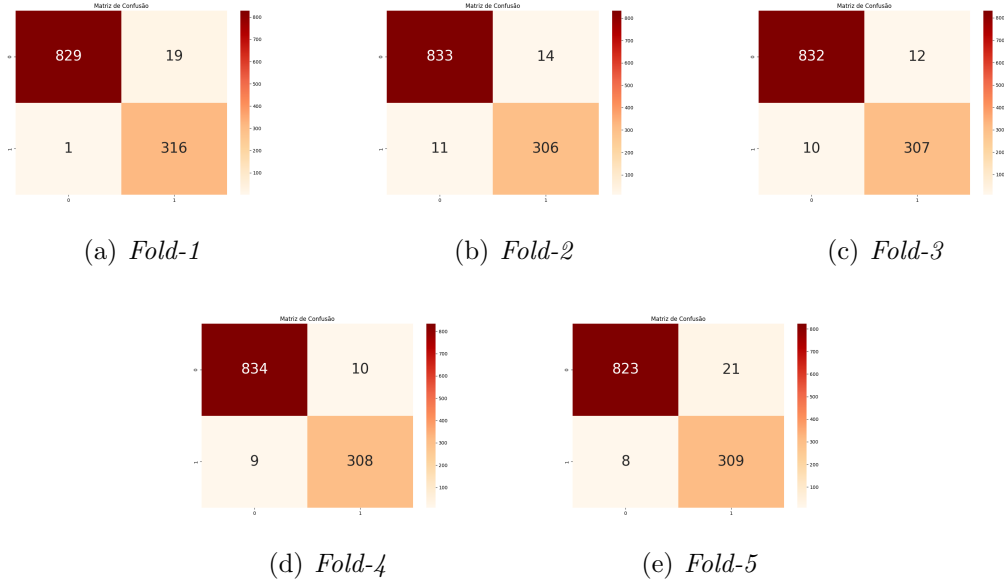


Figura D.6: Matrizes de confusão para *normal*.

D.3 Esclerose Pura sem Crescente

D.3.1 Curvas da Função de Custo

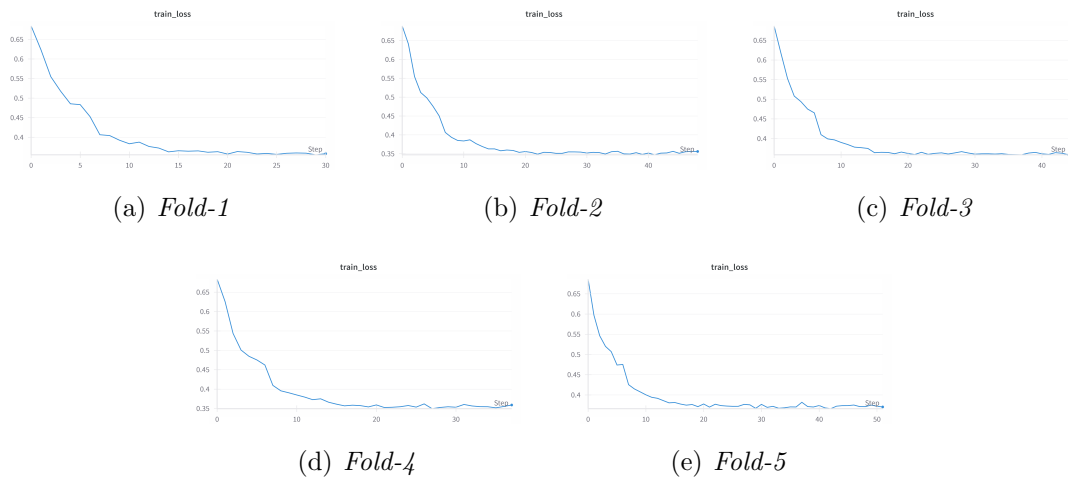


Figura D.7: Curvas da função de custo para *esclerose pura sem crescente*.

D.3.2 Curvas precision-recall

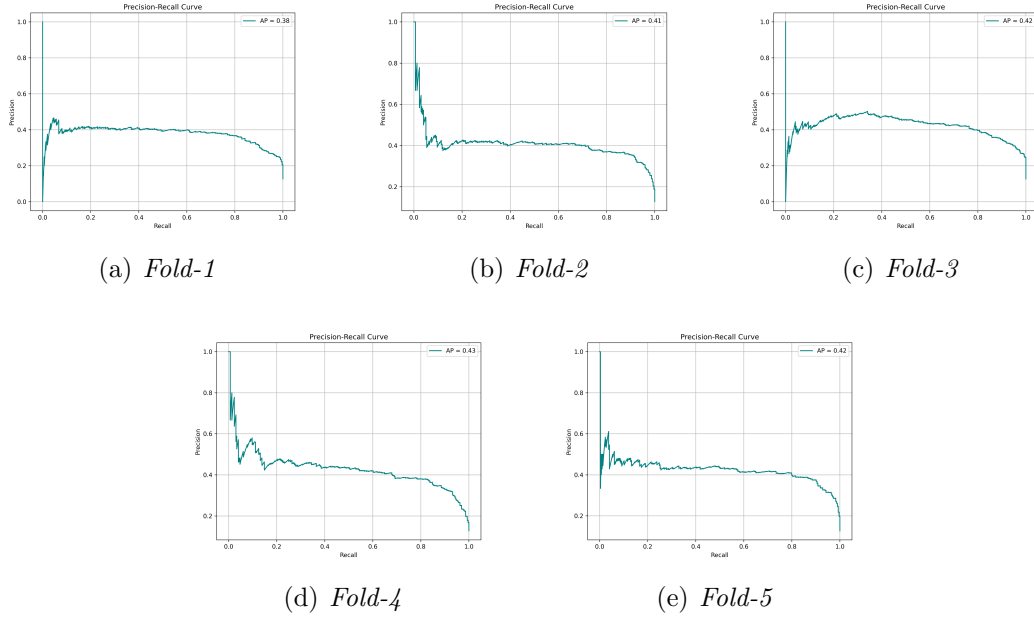


Figura D.8: Curvas *precision-recall* para *esclerose pura sem crescente*.

D.3.3 Matrizes de Confusão

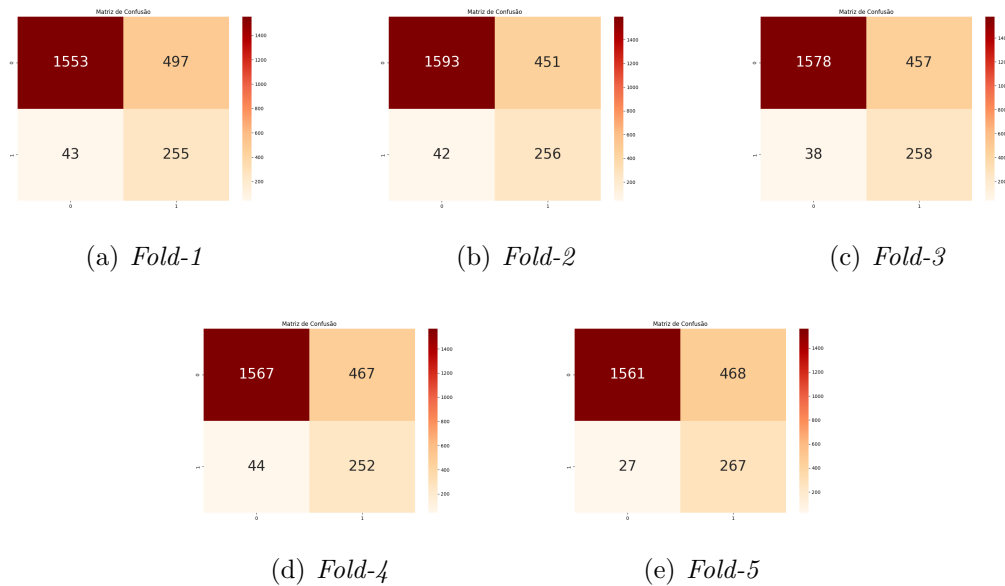


Figura D.9: Matrizes de confusão para *esclerose pura sem crescente*.

D.4 Hipercelularidade

D.4.1 Curvas da Função de Custo

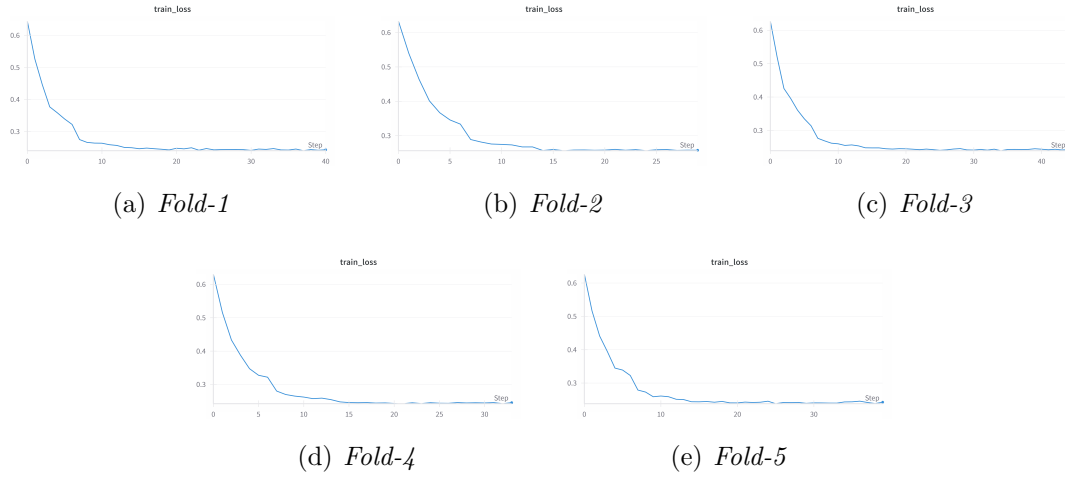


Figura D.10: Curvas da função de custo para *hipercelularidade*.

D.4.2 Curvas precision-recall

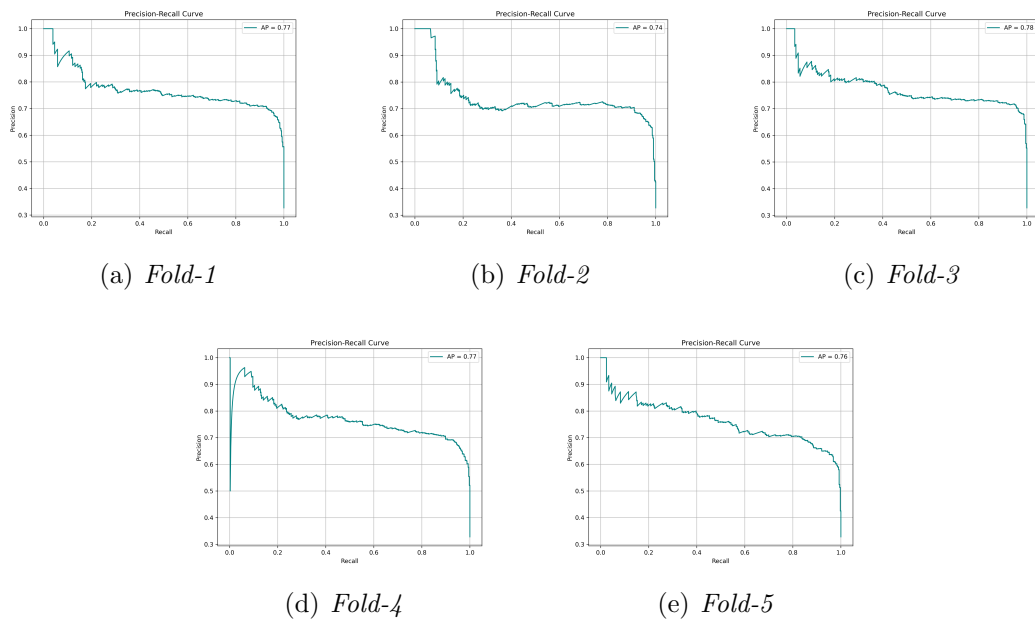


Figura D.11: Curvas *precision-recall* para *hipercelularidade*.

D.4.3 Matrizes de Confusão

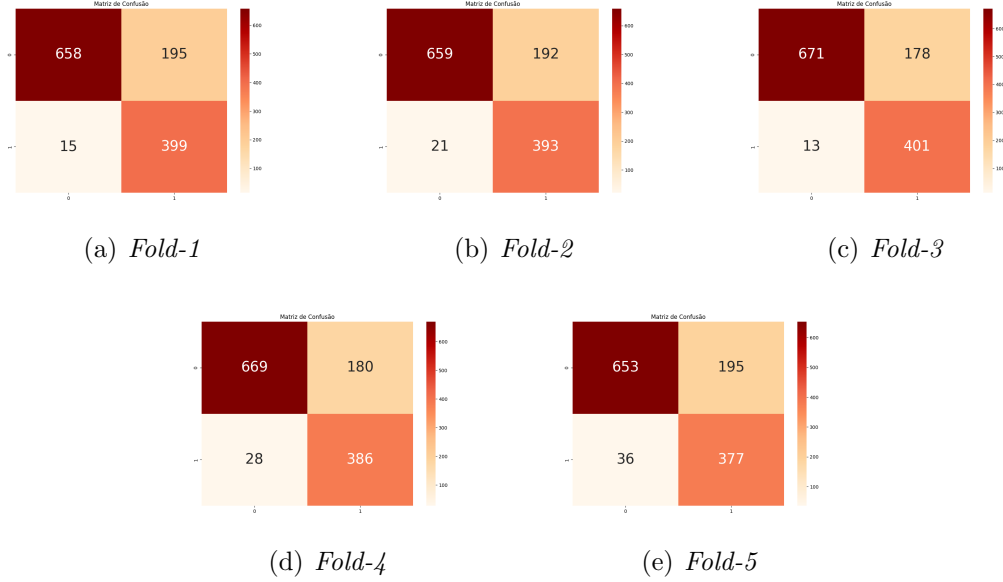


Figura D.12: Matrizes de confusão para *hipercelularidade*.

D.5 Hipercelularidade Pura sem Crescente

D.5.1 Curvas da Função de Custo

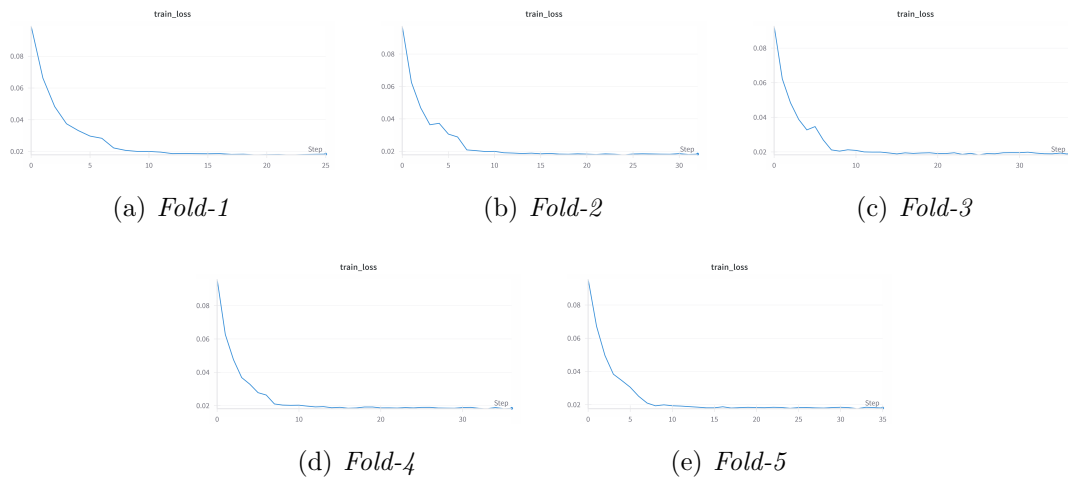


Figura D.13: Curvas da função de custo para *hipercelularidade pura sem crescente*.

D.5.2 Curvas precision-recall

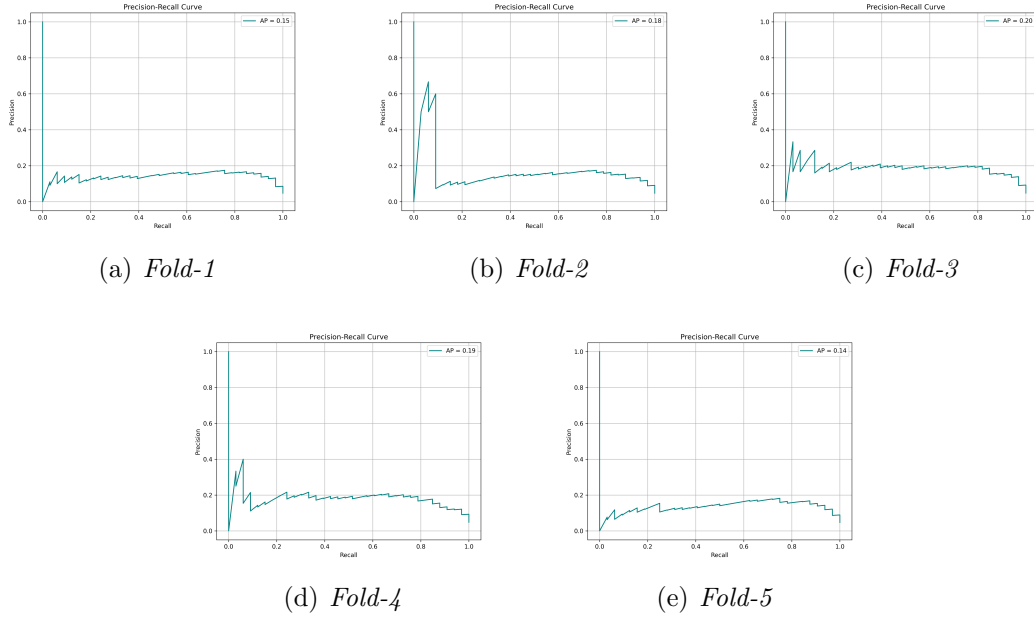


Figura D.14: Curvas *precision-recall* para *hipercelularidade pura sem crescente*.

D.5.3 Matrizes de Confusão

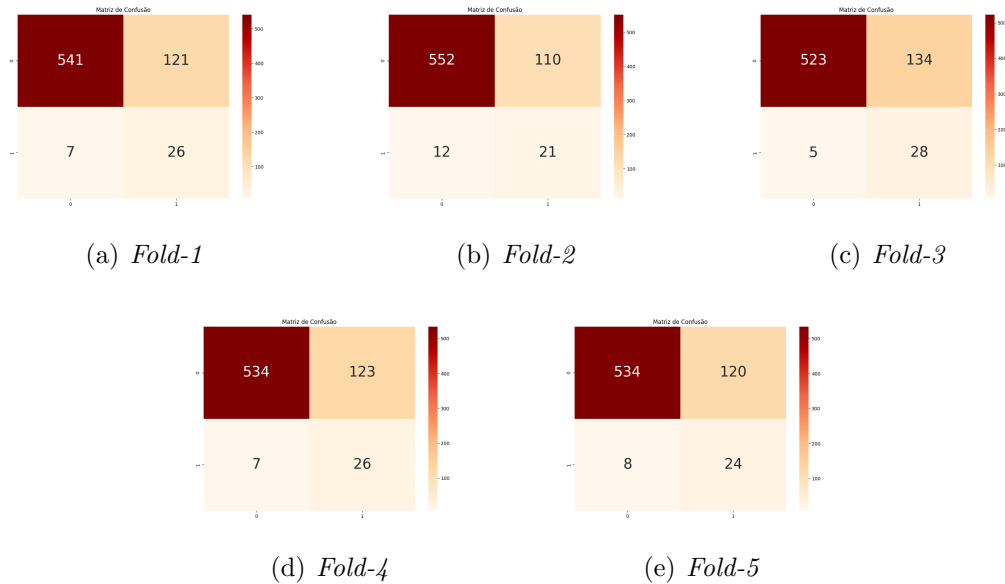


Figura D.15: Matrizes de confusão para *hipercelularidade pura sem crescente*.

D.6 Glomerulonefrite Crescente

D.6.1 Curvas da Função de Custo

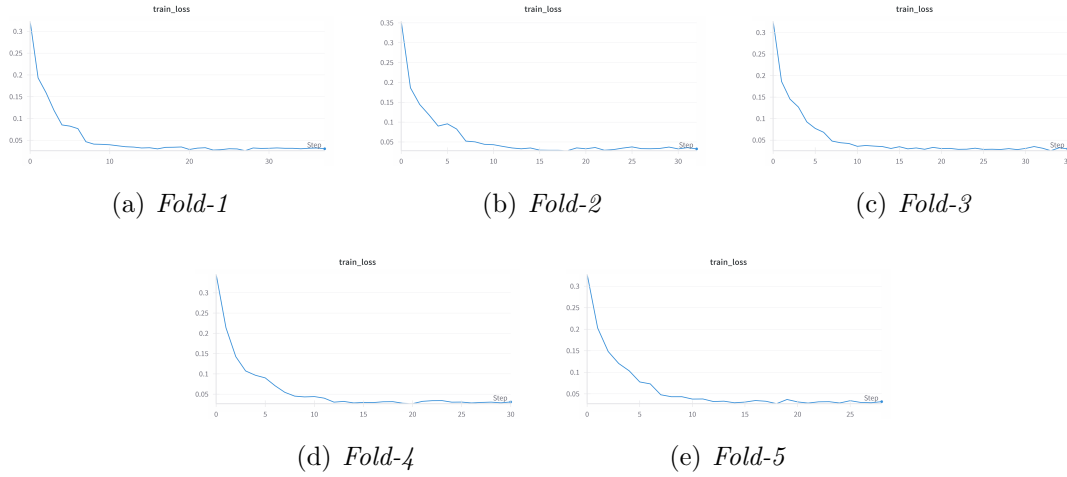


Figura D.16: Curvas da função de custo para *glomerulonefrite crescente*.

D.6.2 Curvas precision-recall

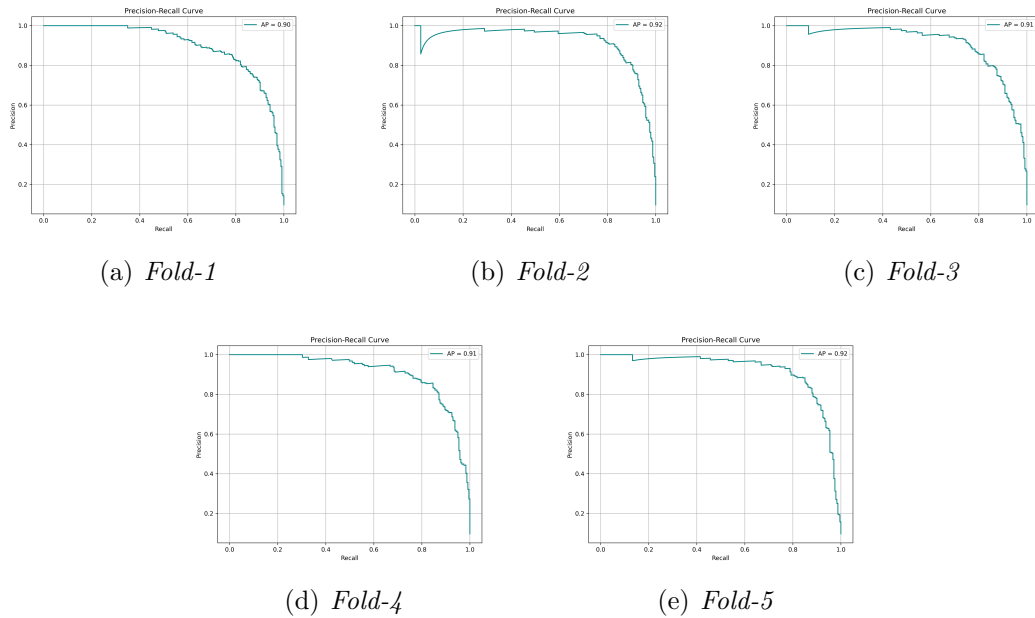


Figura D.17: Curvas *precision-recall* para *glomerulonefrite crescente*.

D.6.3 Matrizes de Confusão

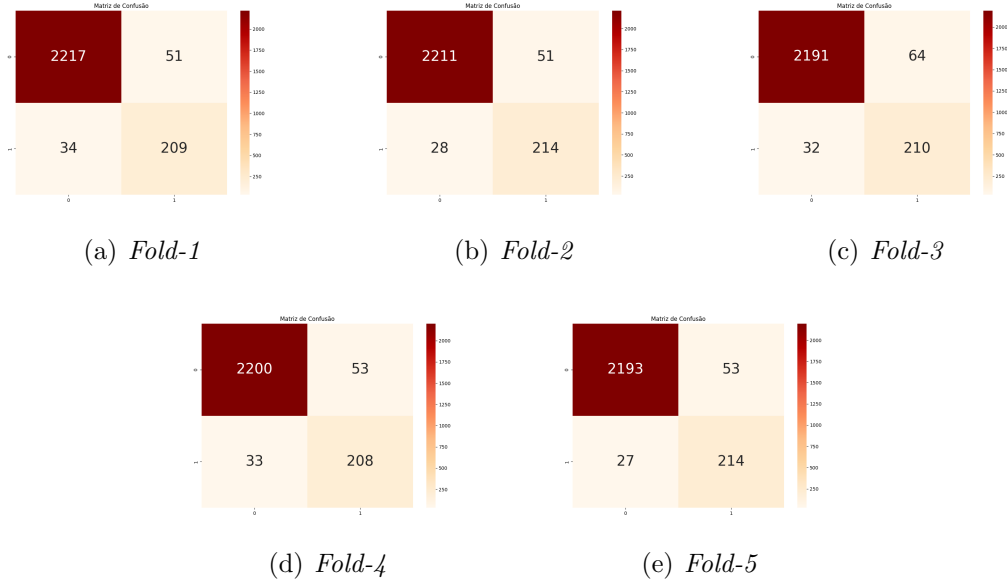


Figura D.18: Matrizes de confusão para *glomerulonefrite crescente*.

D.7 Nefropatia Membranosa

D.7.1 Curvas da Função de Custo

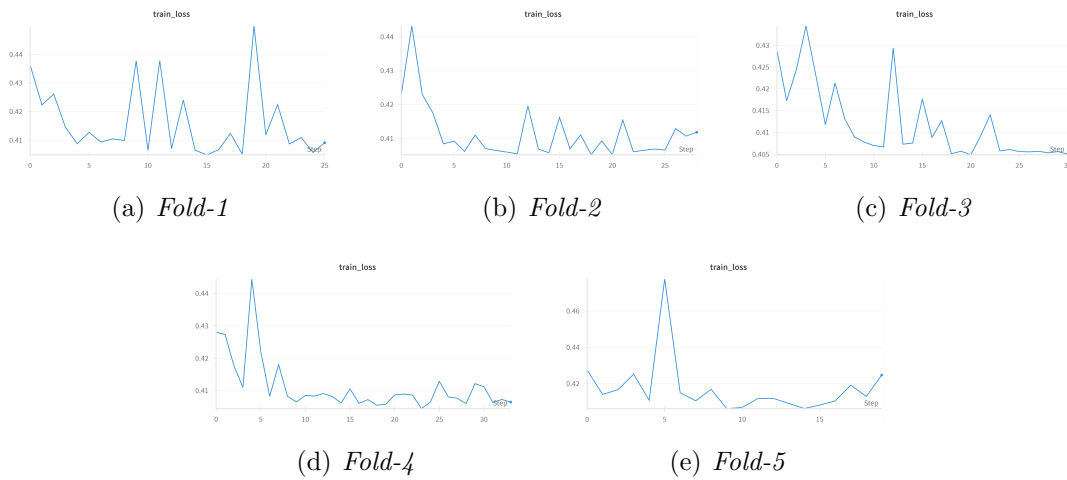


Figura D.19: Curvas da função de custo para *nefropatia membranosa*.

D.7.2 Curvas precision-recall

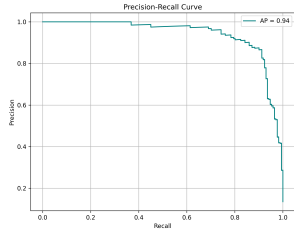
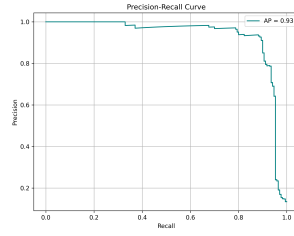
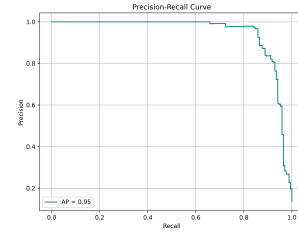
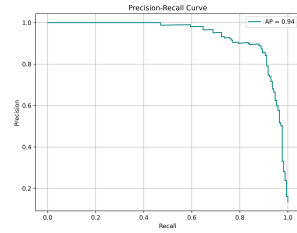
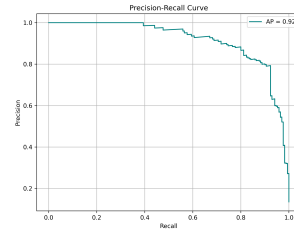
(a) *Fold-1*(b) *Fold-2*(c) *Fold-3*(d) *Fold-4*(e) *Fold-5*

Figura D.20: Curvas *precision-recall* para *nefropatia membranosa*.

D.7.3 Matrizes de Confusão

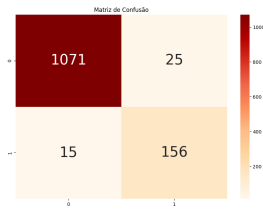
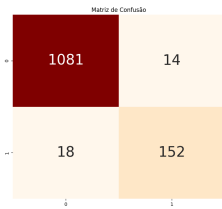
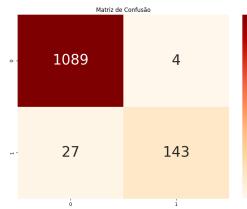
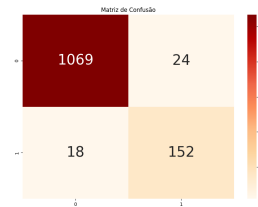
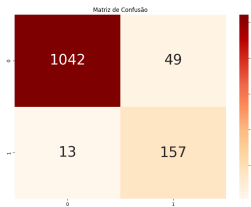
(a) *Fold-1*(b) *Fold-2*(c) *Fold-3*(d) *Fold-4*(e) *Fold-5*

Figura D.21: Matrizes de confusão para *nefropatia membranosa*.

D.8 Esclerose

D.8.1 Curvas da Função de Custo

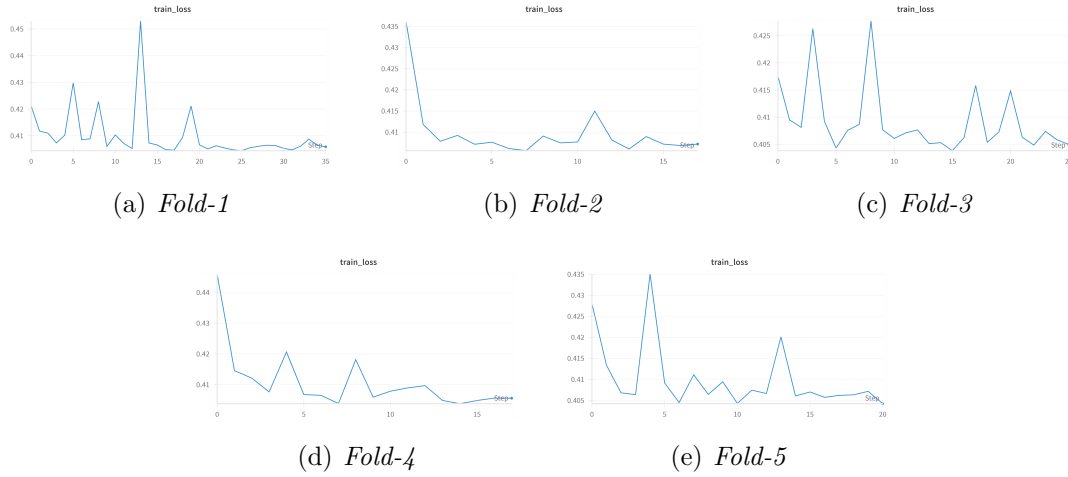


Figura D.22: Curvas da função de custo para *esclerose*.

D.8.2 Curvas precision-recall

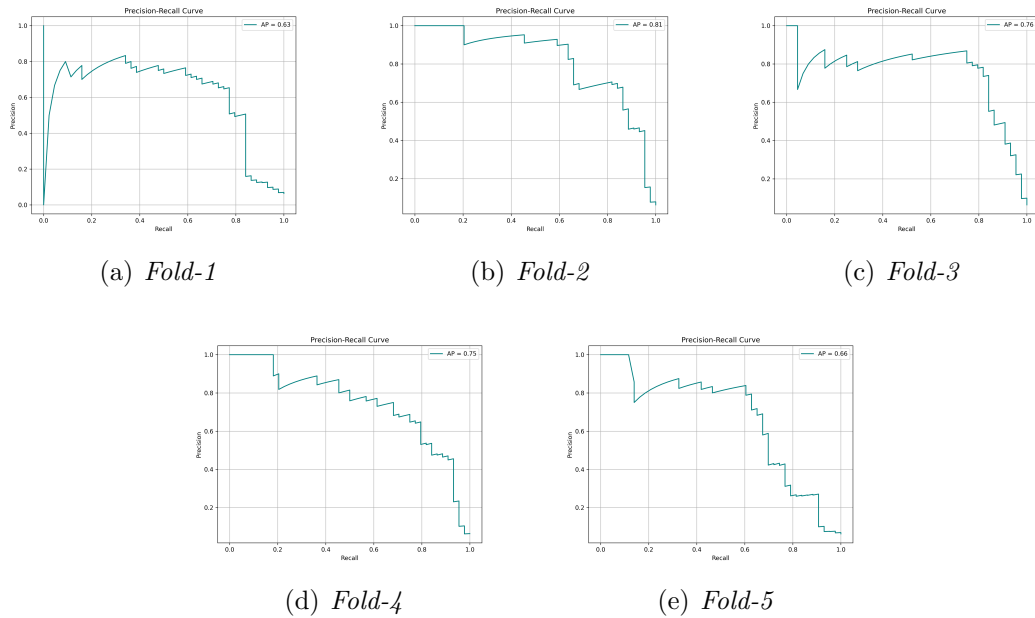


Figura D.23: Curvas *precision-recall* para *esclerose*.

D.8.3 Matrizes de Confusão

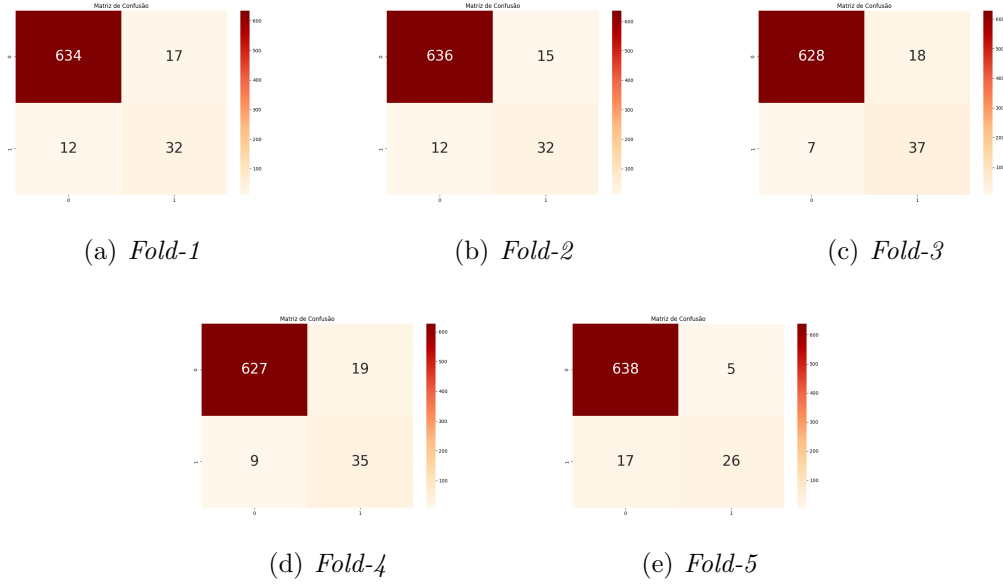


Figura D.24: Matrizes de confusão para *esclerose*.

D.9 Podocitopatia

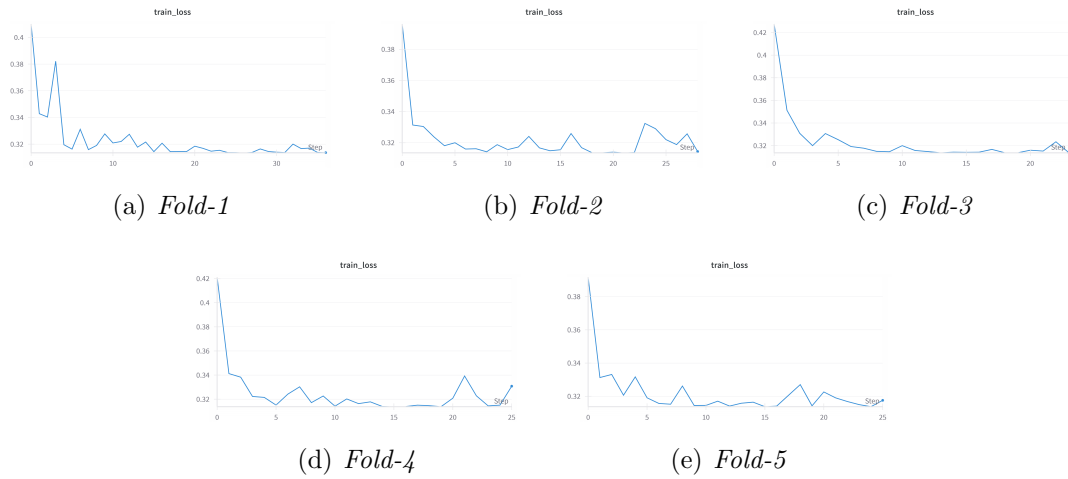


Figura D.25: Curvas da função de custo para *podocitopatia*.

D.9.1 Curvas precision-recall

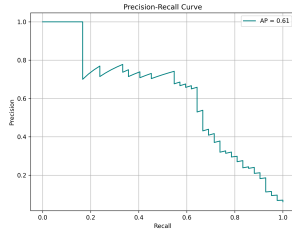
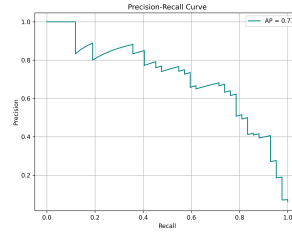
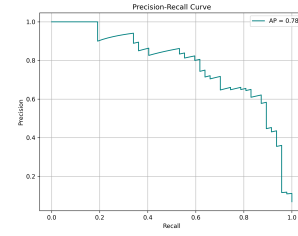
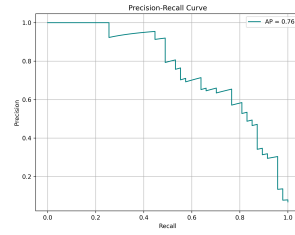
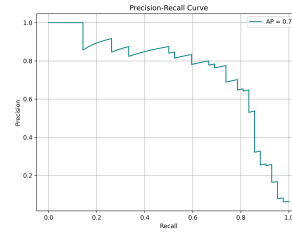
(a) *Fold-1*(b) *Fold-2*(c) *Fold-3*(d) *Fold-4*(e) *Fold-5*

Figura D.26: Curvas *precision-recall* para *podocitopatia*.

D.9.2 Matrizes de Confusão

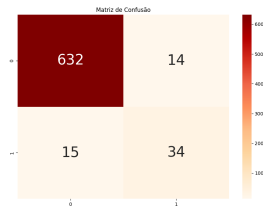
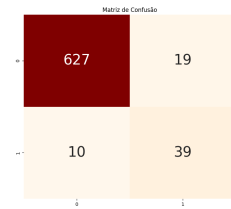
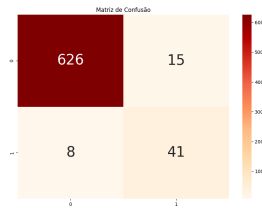
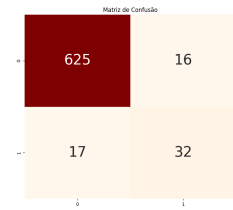
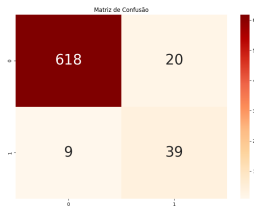
(a) *Fold-1*(b) *Fold-2*(c) *Fold-3*(d) *Fold-4*(e) *Fold-5*

Figura D.27: Matrizes de confusão para *podocitopatia*.