

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ
CÂMPUS CORNÉLIO PROCÓPIO
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

DANIEL HENRIQUE ACORSI ALVES

**ABORDAGENS DE APRENDIZADO ATIVO E PROFUNDO PARA
SÍNTESE E CLASSIFICAÇÃO DE IMAGENS DE
ENTEROPARASITOS**

DISSERTAÇÃO – MESTRADO

CORNÉLIO PROCÓPIO

2019

DANIEL HENRIQUE ACORSI ALVES

**ABORDAGENS DE APRENDIZADO ATIVO E PROFUNDO PARA
SÍNTESE E CLASSIFICAÇÃO DE IMAGENS DE
ENTEROPARASITOS**

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Informática da Universidade Tecnológica Federal do Paraná – UTFPR como requisito para a obtenção do título de “Mestre em Informática”.

Orientadora: Profa. Dra. Priscila Tiemi Maeda Saito

Co-orientador: Prof. Dr. Claiton de Oliveira

CORNÉLIO PROCÓPIO

2019

Dados Internacionais de Catalogação na Publicação

A474 Alves, Daniel Henrique Acorsi

Abordagens de aprendizado ativo e profundo para síntese e classificação de imagens de enteroparasitos / Daniel Henrique Acorsi Alves. – 2019.
113 f. : il. color. ; 31 cm.

Orientadora: Priscila Tiemi Maeda Saito.

Coorientador: Claiton de Oliveira.

Dissertação (Mestrado) – Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Informática, Cornélio Procópio, 2019.

Bibliografia: p. 87-94.

1. Aprendizagem ativa. 2. Processamento de imagens. 3. Helminto. 4. Informática – Dissertações. I. Saito, Priscila Tiemi Maeda, orient. II. Oliveira, Claiton de, coorient. III. Universidade Tecnológica Federal do Paraná. Programa de Pós-Graduação em Informática. IV. Título.

CDD (22. ed.) 004

Biblioteca da UTFPR - Câmpus Cornélio Procópio

Bibliotecário/Documentalista responsável:
Romeu Righetti de Araujo – CRB-9/1676



Título da Dissertação Nº 60:

“ABORDAGENS DE APRENDIZADO ATIVO E PROFUNDO PARA SÍNTESE E CLASSIFICAÇÃO DE IMAGENS DE ENTEROPARASITOS”.

por

Daniel Henrique Acorsi Alves

Orientadora: **Profa. Dra. Priscila Tiemi Maeda Saito**

Co-orientador: **Prof. Dr. Claiton de Oliveira**

Esta dissertação foi apresentada como requisito parcial à obtenção do grau de MESTRE EM INFORMÁTICA – Área de Concentração: Computação Aplicada, pelo Programa de Pós-Graduação em Informática – PPGI – da Universidade Tecnológica Federal do Paraná – UTFPR – Câmpus Cornélio Procópio, às 14h30 do dia 09 de julho de 2019. O trabalho foi _____ pela Banca Examinadora, composta pelos professores:

Profa. Dra. Priscila Tiemi Maeda Saito
(Presidente – UTFPR-CP)

Prof. Dr. Fabrício Martins Lopes
(UTFPR-CP)

Prof. Dr. João Paulo Papa
(UNESP)
Participação à distância via _____

Visto da coordenação:

Danilo Sipoli Sanches
Coordenador do Programa de Pós-Graduação em Informática
UTFPR Câmpus Cornélio Procópio

A Folha de Aprovação assinada encontra-se na Coordenação do Programa.

Av. Alberto Carazzai, 1640 - 86.300-000- Cornélio Procópio – PR.

Tel. +55 (43) 3520-4055 / e-mail: ppgi-cp@utfpr.edu.br / www.utfpr.edu.br/cornelioprocopio/ppgi

AGRADECIMENTOS

Primeiramente agradeço à Deus por permitir que eu concluísse mais essa etapa da minha vida.

Agradeço aos meus pais Marcos Roberto Alves e Celia Acorsi Alves que sempre me apoiaram e incentivaram a dar continuidade aos estudos.

Agradeço à minha esposa Michelle Andrade Klaiber pela paciência, dedicação e total suporte durante essa jornada.

Agradeço imensamente à minha orientadora Profa. Dra. Priscila Tiemi Maeda Saito pela dedicação e incontáveis ajudas em todas as etapas do meu mestrado, obrigado também pelas oportunidades e pela confiança.

Também agradeço ao meu coorientador Prof. Doutor Claiton de Oliveira que sempre esteve disponível para auxiliar e colaborar em diversas etapas do meu mestrado.

Agradeço igualmente ao Prof. Doutor Pedro Henrique Bugatti que contribuiu em diversos momentos e sempre esteve presente durante o desenvolvimento dessa trabalho.

Agradeço aos colegas da ATI - UEL por possibilitarem que eu me ausentasse do trabalho em alguns momentos para dedicação ao mestrado.

Agradeço a todos que, de alguma forma, contribuíram para realização e conclusão dessa dissertação.

Agradeço o suporte dos órgãos de fomento a pesquisa: CNPq (grants #431668/2016-7, #422811/2016-5) CAPES, Fundação Araucária, FAPESP (processo 2014/12236-1), SETI e UTFPR.

RESUMO

ALVES, Daniel Henrique Acorsi. ABORDAGENS DE APRENDIZADO ATIVO E PROFUNDO PARA SÍNTESE E CLASSIFICAÇÃO DE IMAGENS DE ENTEROPARASITOS. 114 f. Dissertação – Mestrado – Programa de Pós-graduação em Informática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2019.

Avanços significativos na área de visão computacional têm sido impulsionados principalmente pelo uso de técnicas de aprendizado profundo por meio das redes neurais convolucionais. No entanto, uma das principais dificuldades na utilização de técnicas de aprendizado profundo é a disponibilidade de dados rotulados, indispensáveis quando se lida com aprendizado supervisionado para a tarefa de classificação. Outros problemas a serem considerados referem-se à: i-) escassez de dados, por exemplo, de imagens, dada a dificuldade de aquisição das mesmas em algumas aplicações; ii-) dificuldade de obtenção de imagens rotuladas por especialistas. Dessa forma, torna-se fundamental o desenvolvimento de mecanismos para obtenção de um conjunto maior de imagens, bem como seus respectivos rótulos. Algumas aplicações podem até apresentar uma quantidade razoável ou até mesmo grandes quantidades de imagens disponíveis, bem como alguns esforços têm sido realizados na tentativa de solucionar os problemas mencionados. No entanto, muitas das amostras do conjunto de imagens podem não ser representativas para o aprendizado, de acordo com o domínio de aplicação, bem como muitas das amostras consideradas no aprendizado podem ser redundantes e desnecessárias, impactando negativamente no desempenho do classificador. Além disso, quando se lida com um grande conjunto de dados e aplicações que exigem tempos de resposta interativos, o processo de classificação pode tornar-se ineficiente e inviável de ser realizado. Portanto, o presente trabalho apresenta a proposta de uma nova abordagem de aprendizado, de forma a obter sintetizadores de imagens e classificadores de padrões mais robustos. Para tanto, integra estratégias de aprendizado ativo nos processos de síntese (por meio do uso de *Generative Adversarial Networks*) e de classificação de imagens, gerando e selecionando imagens sintéticas informativas para o aprendizado do classificador. Além disso, investiga e propõe novas estratégias de aprendizado ativo, de forma a selecionar amostras mais informativas (baseadas em critérios de diversidade e incerteza). Para validação da abordagem de aprendizado proposta, experimentos foram realizados utilizando dados de enteroparasitos. Os resultados obtidos demonstram que a inclusão de estratégias de aprendizado ativo em diferentes etapas da abordagem proposta mostra-se promissora. As estratégias de aprendizado ativo possibilitam a seleção de amostras mais informativas para ambos os processos de síntese e de classificação das imagens.

Palavras-chave: Aprendizado Ativo, Aprendizado Profundo, Processamento e Análise de Imagens, Classificação, Enteroparasites

ABSTRACT

ALVES, Daniel Henrique Acorsi. ACTIVE AND DEEP LEARNING APPROACHES TO SYNTHESIS AND CLASSIFICATION OF ENTEROPARASITES IMAGES. 114 f. Dissertação – Mestrado – Programa de Pós-graduação em Informática, Universidade Tecnológica Federal do Paraná. Cornélio Procópio, 2019.

Significant advances in the computer vision area have been boosted mainly by the use of deep learning techniques through convolutional neural networks. However, one of the main difficulties in using deep learning techniques is the availability of labeled data, which is indispensable when dealing with supervised learning for the task of classification. Other problems to be considered refer to: i-) scarcity of images, for example, images, given the difficulty of acquiring them in some applications; ii-) difficulty in obtaining images labeled by specialists. Thus, it becomes fundamental to develop mechanisms to obtain a larger set of images, as well as that they are labeled. Some applications may even present a reasonable amount or even large amounts of available images, as well as some efforts have been made in attempting to solve the mentioned problems. However, many of the image set samples may not be representative for learning, according to the application domain, and many of the samples considered in the learning can be redundant and unnecessary, negatively impacting the classifier performance. In addition, when dealing with a large set of data and applications that require interactive response times, the classification process may become inefficient and unfeasible to perform. Therefore, the present study presents the proposal for a new learning approach in order to obtain images synthesizers and more robust classifiers. In order to do so, it integrates active learning strategies in the synthesis processes (through the use of Generative Adversarial Networks) and images classification, generating and selecting informative synthetic images for classifier learning. In addition, it investigates and proposes new active learning strategies in order to select more informative samples (based on diversity and uncertainty criteria). To validate the proposed learning approach, experiments were performed using enteroparasites data. The results show that the inclusion of active learning strategies in different stages of the proposed approach is promising. Active learning strategies make it possible to select more informative samples for both image synthesis and classification processes.

Keywords: Active Learning, Deep Learning, Image Processing and Analysis, Classification, Enteroparasites

LISTA DE FIGURAS

FIGURA 1	– Taxa de erro dos algoritmos vencedores na competição ILSVRC de 2010 a 2016.	15
FIGURA 2	– <i>Pipeline</i> do aprendizado de máquina supervisionado.	19
FIGURA 3	– Processo de aprendizado ativo tradicional.	20
FIGURA 4	– Representação de uma Rede Neural Simples (esquerda) e uma Rede Neural Profunda com múltiplas camadas ocultas (direita).	25
FIGURA 5	– Representação da operação de convolução. O elemento central do filtro é colocado sobre o pixel alvo, o qual é substituído pela soma ponderada do mesmo e seus vizinhos.	26
FIGURA 6	– Representação da operação de <i>max pooling</i> com filtro de tamanho 2×2 e tamanho do passo 2.	27
FIGURA 7	– Arquitetura LeNet-5 composta por 2 camadas de convolução com subamostragem e 2 camadas totalmente conectadas.	27
FIGURA 8	– Ilustração da técnica de regularização <i>Dropout</i>	28
FIGURA 9	– Principais funções de ativação para as redes neurais.	28
FIGURA 10	– Arquitetura AlexNet, composta por 5 camadas de convolução e 3 camadas totalmente conectadas.	29
FIGURA 11	– Arquitetura <i>GoogleNet</i> ou <i>Inception-v1</i> composta por 22 camadas, vencedora da tarefa de classificação na competição ILSVRC de 2014 com uma taxa de erro de 6,7%.	30
FIGURA 12	– Taxa de erro durante o treinamento (esquerda) e taxa de erro durante o teste (direita).	31
FIGURA 13	– Bloco de conexão residual da arquitetura ResNet.	31
FIGURA 14	– Modelo das Generative Adversarial Networks (GANs).	32
FIGURA 15	– Arquitetura da DCGAN.	34
FIGURA 16	– <i>Transfer Learning</i> aplicado à tarefa de classificação no contexto de redes neurais.	37
FIGURA 17	– Exemplos de imagens e respectivas classes do conjunto de dados CIFAR-10.	40
FIGURA 18	– Exemplos de imagens do subconjunto de dados PI@ntLeaves II.	41
FIGURA 19	– Distribuição das imagens em cada classe do subconjunto de dados PI@ntLeaves II.	41
FIGURA 20	– Exemplos de imagens do conjunto de enteroparasitos. (a) espécies de enteroparasitos. (b) impurezas fecais similares às espécies de enteroparasitos.	42
FIGURA 21	– <i>Pipeline</i> da análise comparativa de desempenho, considerando abordagens de aprendizados tradicionais e de aprendizado profundo para classificação de folhas de plantas.	44
FIGURA 22	– Arquitetura <i>Inception-v3</i> , destacando a saída da camada de <i>dropout</i> utilizada como vetor de características obtido por G_1	45
FIGURA 23	– Acurácias médias obtidas com RF a partir das reduções de dimensionalidade das características (Inception-v3) utilizando PCA para cada conjunto de dados (LEA-53, CIFAR-10, PAR-15 e PAR-16).	50

FIGURA 24	– Amostras obtidas na primeira iteração pelas estratégias EN-CLU, LC-CLU e MS-CLU.	52
FIGURA 25	– Etapa de organização da estratégia de aprendizado ativo <i>Minimum-Spanning Tree Cluster Decreasing Disagree Edges</i> (MST-CLU-DDE).	53
FIGURA 26	– Etapa de organização da estratégia de aprendizado ativo <i>Minimum-Spanning Tree Cluster Decreasing Sampling</i> (MST-CLU-DS).	56
FIGURA 27	– Etapa de organização da estratégia de aprendizado ativo <i>Minimum-Spanning Tree Cluster Root Disagree Sampling</i> (MST-CLU-RDS).	59
FIGURA 28	– Médias das acurácias obtidas sem o uso de estratégias de aprendizado ativo com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	62
FIGURA 29	– Médias das acurácias obtidas pelas estratégias de aprendizado ativo EN e EN-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2. . .	63
FIGURA 30	– Médias das acurácias obtidas pelas estratégias de aprendizado ativo MS e MS-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2. . .	64
FIGURA 31	– Médias das acurácias obtidas pelas estratégias de aprendizado ativo LC e LC-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2. . .	65
FIGURA 32	– Médias das acurácias obtidas pelas estratégias de aprendizado ativo EN-CLU, MS-CLU e LC-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	67
FIGURA 33	– Médias das acurácias obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	68
FIGURA 34	– Médias de acurácias e de tempos de organização, seleção e classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	70
FIGURA 35	– Médias das acurácias obtidas pela estratégia de aprendizado ativo MS-CLU com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	71
FIGURA 36	– <i>Pipeline</i> da abordagem de aprendizado proposta para anotação de imagens, envolvendo a aplicação de estratégias de aprendizado ativo na construção de bases de imagens sintéticas e no aprendizado de classificadores.	73
FIGURA 37	– Fluxo para análise das métricas utilizadas para avaliação das imagens geradas pelas GANs.	76
FIGURA 38	– Exemplos de imagens (reais e sintéticas) do conjunto de enteroparasitos. As cinco primeiras imagens referem-se às imagens reais e as cinco imagens seguintes correspondem às imagens sintéticas geradas pela BEGAN.	77
FIGURA 39	– Fluxos do <i>framework</i> proposto.	82
FIGURA 40	– Médias das acurácias obtidas pelas estratégias de aprendizado ativo DBE,	

	MS-CLU, MST-BE e RBE com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	96
FIGURA 41	– Médias de acurácias, tempo de organização, tempo de seleção e tempo de classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador k -NN nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	98
FIGURA 42	– Médias de acurácias, tempo de organização, tempo de seleção e tempo de classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador RF nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	100
FIGURA 43	– Médias de acurácias, tempo de organização, tempo de seleção e tempo de classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	102
FIGURA 44	– Médias das acurácias obtidas pelas estratégia de aprendizado ativo MST-CLU-DDE com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	104
FIGURA 45	– Médias das acurácias obtidas pelas estratégia de aprendizado ativo MST-CLU-DS com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	106
FIGURA 46	– Médias das acurácias obtidas pelas estratégia de aprendizado ativo MST-CLU-RDS com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	108
FIGURA 47	– Médias das acurácias obtidas pelas estratégia de aprendizado ativo RDS com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.	110

LISTA DE TABELAS

TABELA 1	– Quantidades de classes e imagens em cada um dos subconjuntos de parasitos e impurezas intestinais.	42
TABELA 2	– Descrição das classes e distribuição das imagens em cada classe dos subconjuntos de parasitos e impurezas intestinais.	43
TABELA 3	– Grupos de técnicas de extração de características considerados nos experimentos e as respectivas quantidades de características obtidas.	45
TABELA 4	– Técnicas de seleção e quantidade de características obtidas para cada extrator.	46
TABELA 5	– Acurácias médias e desvios-padrão obtidos no conjunto Pl@ntLeaves II sem seleção de características. Os valores sublinhados representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.	48
TABELA 6	– Acurácias médias e desvios-padrão obtidos no conjunto Pl@ntLeaves II com seleção das melhores características. Os valores sublinhados representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.	48
TABELA 7	– Acurácias médias e desvios-padrão obtidos no conjunto Pl@ntLeaves II combinando as melhores características de G_1 . Os valores sublinhados representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.	48
TABELA 8	– Acurácias médias e desvios-padrão obtidos pelas arquiteturas <i>ResNet18</i> e <i>Inception-v3</i> nos conjuntos de dados de enteroparasitos.	49
TABELA 9	– Métricas <i>Inception Score</i> (IS), <i>Mode Score</i> (MD) e <i>Fréchet Inception Distance</i> (FID) para avaliação das imagens (de cada classe) de enteroparasitos geradas pela arquitetura BEGAN. São destacados os melhores (negrito) e os piores (underline) resultados obtidos por cada métrica.	76
TABELA 10	– Acurácias (%) obtidas pelo classificador SVM, para cada conjunto de enteroparasitos (LAR-2, EGG-8, EGG-9, PRO-6, PRO-7, PAR-15 e PAR-16), considerando as abordagens <i>baseline</i> (contendo apenas amostras reais) e as arquiteturas GAN, DCGAN, WGAN, WGAN-GP e BEGAN (utilizando imagens sintéticas e as métricas <i>GAN-train</i> e <i>GAN-test</i>).	79
TABELA 11	– Resultados obtidos utilizando imagens reais (baseline) e imagens sintéticas por meio das arquiteturas GAN, DCGAN, WGAN, WGAN-GP e BEGAN, considerando os valores das métricas IS, MD, FID, <i>GAN-train</i> e <i>GAN-test</i> no conjunto de dados CIFAR-10. <i>GAN-train</i> e <i>GAN-test</i> são medidas de acurácia com o classificador SVM.	80
TABELA 12	– Acurácias obtidas pela estratégia MS-CLU, utilizando o classificador SVM para cada fluxo de operação do <i>framework</i> proposto (GAN_Flow ,	

	AL-GAN_Flow, GAN-AL_Flow e AL-GAN-AL_Flow) e para cada conjunto de enteroparasitos.	82
TABELA 13	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto LEA-53.	111
TABELA 14	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PAR-16.	111
TABELA 15	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PAR-15.	111
TABELA 16	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto EGG-9.	112
TABELA 17	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto EGG-8.	112
TABELA 18	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PRO-6.	112
TABELA 19	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PRO-7.	112
TABELA 20	– Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto LAR-2.	113

LISTA DE SIGLAS

BEGAN	Boundary Equilibrium Generative Adversarial Network
BN	Batch Normalization
CFS	Correlation-based Feature Selection
CNNs	Convolutional Neural Networks
DBE	Decreasing Boundary Edges
DCGAN	Deep Convolutional Generative Adversarial Network
FID	Fréchet Inception Distance
GANs	Generative Adversarial Networks
IGR	Information Gain Ratio
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IS	Inception Score
k -NN	k -Nearest Neighbor
LSTM	Long Short Term Memory
MST	Minimum-Spanning Tree
MST-BE	Minimum-Spanning Tree Boundary Edges
NB	Naive Bayes
OPF	Optimum-Path Forest
PCA	Principal Components Analysis
RBE	Random Boundary Edges
RDS	Root-Distance Based Sampling
RF	Random Forest
SOM	Self-Organizing Map
SVM	Support Vector Machine
UC	Unsupervised Clustering
WGAN	Wasserstein GAN
WGAN-GP	Wasserstein GAN Gradient Penalty

SUMÁRIO

1	INTRODUÇÃO	14
1.1	JUSTIFICATIVA	15
1.2	OBJETIVOS GERAIS	16
1.2.1	OBJETIVOS ESPECÍFICOS	16
1.3	ORGANIZAÇÃO DO TEXTO	17
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	APRENDIZADO DE MÁQUINA	18
2.2	APRENDIZADO ATIVO	19
2.2.1	Técnicas Clássicas Baseadas em Medidas de Incerteza	20
2.2.2	Unsupervised Clustering (UC)	21
2.2.3	Random Boundary Edges (RBE)	22
2.2.4	Decreasing Boundary Edges (DBE)	22
2.2.5	Minimum-Spanning Tree Boundary Edges (MST-BE)	23
2.2.6	Root-Distance Based Sampling (RDS)	24
2.3	APRENDIZADO PROFUNDO	24
2.3.1	Convolutional Neural Network (CNN)	25
2.3.1.1	AlexNet	27
2.3.1.2	GoogleNet / Inception	29
2.3.1.3	Residual Network (ResNet)	30
2.3.1.4	Além das CNNs	31
2.3.2	Generative Adversarial Network (GAN)	32
2.3.2.1	Deep Convolutional Generative Adversarial Network (DCGAN)	33
2.3.2.2	Boundary Equilibrium Generative Adversarial Network (BEGAN)	34
2.3.2.3	Métricas das GANs	35
2.4	TRANSFER LEARNING	36
2.5	ABORDAGENS DE APRENDIZADO ATIVO E PROFUNDO	37
3	DESCRIÇÃO DOS CONJUNTOS DE DADOS	39
4	CONTRIBUIÇÃO 1 - DESCRITORES PARA IMAGENS - DEEP FEATURES VS HANDCRAFTED FEATURES	44
4.1	METODOLOGIA PROPOSTA	44
4.2	EXPERIMENTOS E RESULTADOS	47
5	CONTRIBUIÇÃO 2 - NOVAS ESTRATÉGIAS DE APRENDIZADO ATIVO	51
5.1	ENTROPY CLUSTER (EN-CLU), LEAST CONFIDENCE CLUSTER (LC-CLU) E MARGIN SAMPLING CLUSTER (MS-CLU)	51
5.2	MINIMUM-SPANNING TREE CLUSTER DECREASING DISAGREE EDGES (MST-CLU-DDE)	52
5.3	MINIMUM-SPANNING TREE CLUSTER DECREASING SAMPLING (MST-CLU-DS)	55
5.4	MINIMUM-SPANNING TREE CLUSTER ROOT DISAGREE SAMPLING (MST-CLU-RDS)	58

5.5 EXPERIMENTOS E RESULTADOS	60
6 CONTRIBUIÇÃO 3 - FRAMEWORK PARA SÍNTESE E CLASSIFICAÇÃO DE	
ENTEROPARASITOS	72
6.1 METODOLOGIA PROPOSTA	72
6.2 ANÁLISE DAS MÉTRICAS DE AVALIAÇÃO DAS GANs	74
6.3 EXPERIMENTOS E RESULTADOS	81
7 CONCLUSÕES	85
REFERÊNCIAS	87
Apêndice A – ACURÁCIAS ESTRATÉGIAS RBE, DBE, MST-BE E MS-CLU COM	
CLASSIFICADOR SVM	95
Apêndice B – DESEMPENHO DAS ESTRATÉGIAS MS-CLU, MST-CLU-DS, MST-	
CLU-DDE, MST-CLU-RDS E RDS COM CLASSIFICADOR K-NN	97
Apêndice C – DESEMPENHO DAS ESTRATÉGIAS MS-CLU, MST-CLU-DS, MST-	
CLU-DDE, MST-CLU-RDS E RDS COM CLASSIFICADOR RF	99
Apêndice D – DESEMPENHO DAS ESTRATÉGIAS MS-CLU, MST-CLU-DS, MST-	
CLU-DDE, MST-CLU-RDS E RDS COM CLASSIFICADOR NB	101
Apêndice E – ACURÁCIAS DA ESTRATÉGIA MST-CLU-DDE COM CLASSIFI-	
CADORES SVM, K-NN, RF E NB	103
Apêndice F – ACURÁCIAS DA ESTRATÉGIA MST-CLU-DS COM CLASSIFI-	
DORES SVM, K-NN, RF E NB	105
Apêndice G – ACURÁCIAS DA ESTRATÉGIA MST-CLU-RDS COM CLASSIFI-	
CADORES SVM, K-NN, RF E NB	107
Apêndice H – ACURÁCIAS DA ESTRATÉGIA RDS COM CLASSIFICADORES	
SVM, K-NN, RF E NB	109
Apêndice I – MÉDIA EM SEGUNDOS DOS TEMPOS DE ORGANIZAÇÃO, SELEÇÃO	
E CLASSIFICAÇÃO OBTIDOS PELAS ESTRATÉGIAS MS-CLU, MST-CLU-	
DDE, MST-CLU-DS, MST-CLU-RDS E RDS COM O CLASSIFICADOR SVM	
EM CADA CONJUNTO DE DADOS	111

1 INTRODUÇÃO

A área de visão computacional tem apresentado avanços expressivos ocasionados principalmente pelo emprego de técnicas e modelos de aprendizado profundo. Um exemplo clássico são os veículos autônomos (BOJARSKI et al., 2016) que utilizam sensores e câmeras para coletar e analisar dados do ambiente ao seu redor em tempo real. Esses dados são processados por um modelo de aprendizado profundo e possibilitam a condução do veículo pelas vias sem a intervenção de um humano.

Outro destaque refere-se ao projeto ImageNet (RUSSAKOVSKY et al., 2015), o qual une esforços para disponibilizar uma grande base de dados de imagens rotuladas. Os dados rotulados são fundamentais na etapa de treinamento de um modelo de aprendizado supervisionado. A ImageNet também promove, desde 2010, a *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC), competição com a finalidade de automatizar os processos de recuperação e rotulação de imagens, cuja principal tarefa é a de classificação das imagens em suas respectivas categorias. Um conjunto de dados de aproximadamente 1,4 milhões de amostras de imagens totalmente rotuladas divididas em 1.000 classes encontra-se disponível. Este conjunto de dados é considerado um *baseline* para avaliação de desempenho de modelos de redes neurais profundas. A Figura 1 ilustra a evolução dos algoritmos vencedores ao longo dos anos.

Em Chandraratne et al. (2007) é proposta a classificação de tipos de cortes de carne de cordeiro usando redes neurais e métodos estatísticos. No trabalho (ABEDINI et al., 2015), foi demonstrado o desempenho de algoritmos de aprendizado de máquina para classificação de imagens médicas. Os autores desenvolveram uma abordagem generalizada, utilizando imagens de diversas partes do corpo humano, demonstrando a robustez do método. Em Burlina et al. (2017), uma rede neural convolucional e técnicas de *transfer learning* são utilizadas para demonstrar a superioridade na detecção de degeneração macular quando comparado a um especialista humano.

Apesar dessas técnicas de aprendizado se mostrarem eficazes, muitas amostras consideradas no aprendizado podem ser redundantes e desnecessárias, além de impactar negativa-

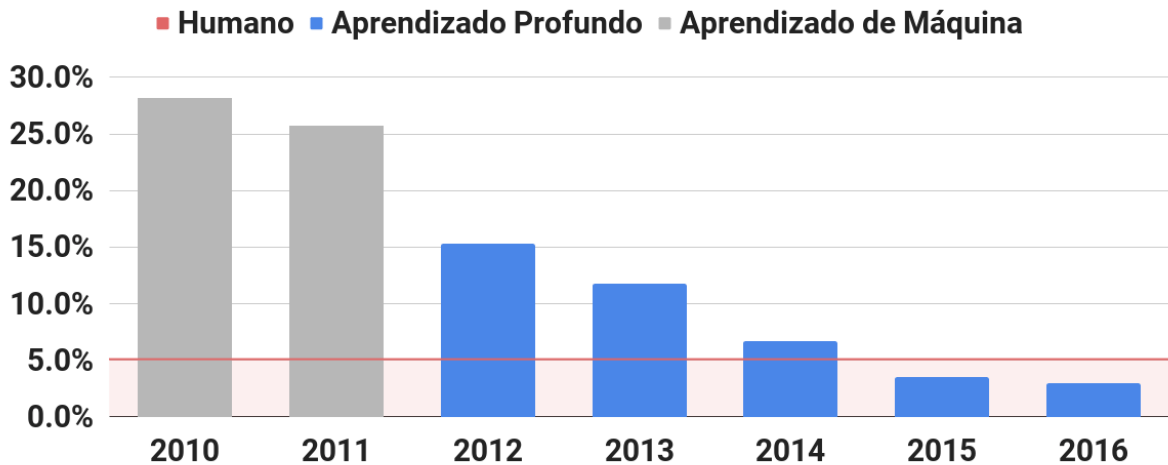


Figura 1: Taxa de erro dos algoritmos vencedores na competição ILSVRC de 2010 a 2016.

mente o desempenho do classificador. Além disso, quando se lida com um grande conjunto de dados e aplicações que exigem tempos de resposta interativos, o processo de classificação pode tornar-se ineficiente e inviável de ser realizado.

Para contornar esses desafios, técnicas de aprendizado ativo têm sido utilizadas, selecionando amostras mais informativas (mais diversas ou incertas) para o aprendizado do classificador. Esforços têm sido realizados e largamente aplicados em diferentes problemas e aplicações reais na tentativa de obter melhores acurácias mais rapidamente, em menos iterações de aprendizado, bem como envolvendo minimamente o especialista no processo de aprendizado.

Portanto, o presente trabalho visa analisar estratégias de aprendizado ativo mais robustas em conjunto com técnicas de aprendizado profundo para melhorar a eficácia e eficiência dos processos de síntese e de classificação de imagens.

1.1 JUSTIFICATIVA

A classificação automática e eficiente de imagens depende, primordialmente, da disponibilidade de amostras devidamente anotadas por especialistas. Essas amostras são submetidas para treinamento de um método de aprendizado supervisionado. No entanto, além da escolha do método adequado, também é importante analisar o impacto que cada amostra apresenta em cada etapa do aprendizado. Técnicas de aprendizado ativo possibilitam a seleção das amostras mais relevantes para o treinamento de um classificador, melhorando o processo como um todo.

Esforços no sentido de criar melhores estratégias de redução, organização e seleção dos dados têm sido despendidos e de certa forma bem sucedidos em diferentes domínios de

aplicação (HUANG et al., 2010; SAITO et al., 2014a, 2014b, 2015; REKER; SCHNEIDER, 2015; SAITO, 2015; GAL et al., 2017; BRESSAN et al., 2018, 2019).

Considerando a escassez de imagens e dada a dificuldade de aquisição em algumas aplicações, aliada com a necessidade de utilização de uma quantidade razoável de imagens para obtenção de elevadas acurácias de classificação, torna-se importante o desenvolvimento de mecanismos para obtenção de mais imagens. Outras aplicações podem apresentar uma quantidade razoável ou até mesmo grandes quantidades de imagens. No entanto, muitas podem não ser representativas para o aprendizado de acordo com o domínio da aplicação, havendo a necessidade de geração de conjuntos de imagens sintéticas específicas do domínio.

Portanto, pretende-se também desenvolver estratégias para a geração e seleção de imagens sintéticas mais informativas para o aprendizado, de forma a obter sintetizadores de imagens, bem como classificadores de padrões mais robustos.

1.2 OBJETIVOS GERAIS

O presente trabalho tem como objetivo principal o estudo, o desenvolvimento e a validação de uma nova abordagem de aprendizado, incluindo estratégias de aprendizado ativo, de forma a selecionar amostras mais informativas, para os processos de síntese e de classificação de imagens de enteroparasitos.

1.2.1 OBJETIVOS ESPECÍFICOS

1. Obtenção dos conjuntos de características que melhor descrevem cada conjunto de imagens, de forma a melhorar os resultados nas etapas de aprendizado posteriores.
2. Proposta e validação de novas estratégias de aprendizado ativo, incluindo alterações e comparações envolvendo outras técnicas da literatura, de forma a obter estratégias mais adequadas a serem consideradas na abordagem proposta.
3. Análise de diferentes arquiteturas de *Generative Adversarial Networks* (GANs) e métricas de avaliação das mesmas para geração de imagens sintéticas, de forma a obter imagens com qualidade adequada para serem consideradas na abordagem proposta.
4. Avaliação do impacto da inclusão de imagens sintéticas no processo de classificação.
5. Avaliação do desempenho de estratégias de aprendizado ativo, na seleção de amostras mais informativas, para os processos de síntese e de classificação das imagens.

6. Desenvolvimento de um *framework* de ponta a ponta para síntese e classificação eficaz e eficiente de imagens.

1.3 ORGANIZAÇÃO DO TEXTO

O presente trabalho apresenta a seguinte organização:

- O Capítulo 2 apresenta a fundamentação teórica desse trabalho, incluindo conceitos de aprendizado de máquina, aprendizado ativo, e as principais técnicas de aprendizado profundo na área de visão computacional, que incluem aplicações na etapa extração de descritores e classificação de imagens (através das arquiteturas de *Convolutional Neural Networks* (CNNs) e na síntese de imagens (através das arquiteturas de GANs).
- No Capítulo 3 são apresentados os conjuntos de dados que são explorados nas diferentes etapas desse trabalho.
- No Capítulo 4 são apresentados os experimentos, análises, resultados e contribuições referentes à escolha de descritores para as imagens.
- No Capítulo 5 são apresentadas novas propostas de estratégias de aprendizado, bem como experimentos, discussões e resultados obtidos.
- O Capítulo 6 descreve o *framework* proposto para classificação de imagens, incluindo experimentos, discussões e resultados.
- No Capítulo 7 são expostas as considerações finais.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta Seção são apresentados conceitos e trabalhos relacionados, incluindo os principais algoritmos de aprendizado de máquina, técnicas de aprendizado ativo, modelos de aprendizado profundo, bem como esforços com a proposta de combinação de aprendizados ativo e profundo encontrados na literatura.

2.1 APRENDIZADO DE MÁQUINA

O aprendizado de máquina é uma área da ciência da computação, na qual o propósito é encontrar de maneira automática um bom preditor a partir de dados de entrada. Um dos pré-requisitos para o emprego das técnicas tradicionais de aprendizado de máquina é a obtenção dos melhores descritores para os dados, sendo que essa etapa é determinante na qualidade do aprendizado de máquina tradicional.

O aprendizado de máquina é tipicamente dividido em três tipos de categorias: i-) aprendizado supervisionado, o qual utiliza dados rotulados para o treinamento; ii-) aprendizado não supervisionado (ou *clustering*), o qual utiliza dados não rotulados para o treinamento e iii-) aprendizado semi-supervisionado, o qual utiliza dados rotulados e não rotulados para treinamento (SHALEV-SHWARTZ; BEN-DAVID, 2014).

Tais abordagens de aprendizado têm sido amplamente utilizadas em diferentes aplicações, tais como em bioinformática com sequenciamento de DNA (ALIPANAHI et al., 2015), diagnósticos médicos (KOUROU et al., 2015), processamento de linguagem natural (CHO et al., 2014), reconhecimento de fala (AMODEI et al., 2016), análise de sentimentos (LIU, 2012), entre outros.

A Figura 2 ilustra o *pipeline* do aprendizado supervisionado de máquina. As técnicas de aprendizado supervisionado utilizam dados previamente rotulados para o treinamento. Diversos algoritmos têm sido propostos, dentre eles, podem ser destacados: *k-Nearest Neighbors* (*k*-NN) (COVER; HART, 1967), *Support Vector Machine* (SVM) (HEARST et al., 1998), *Ran-*

om Forest (RF) (BREIMAN, 2001), *Naive Bayes* (NB) (DOMINGOS; PAZZANI, 1997), C4.5 (QUINLAN, 2014), *Optimum-Path Forest* (OPF) (PAPA et al., 2009).

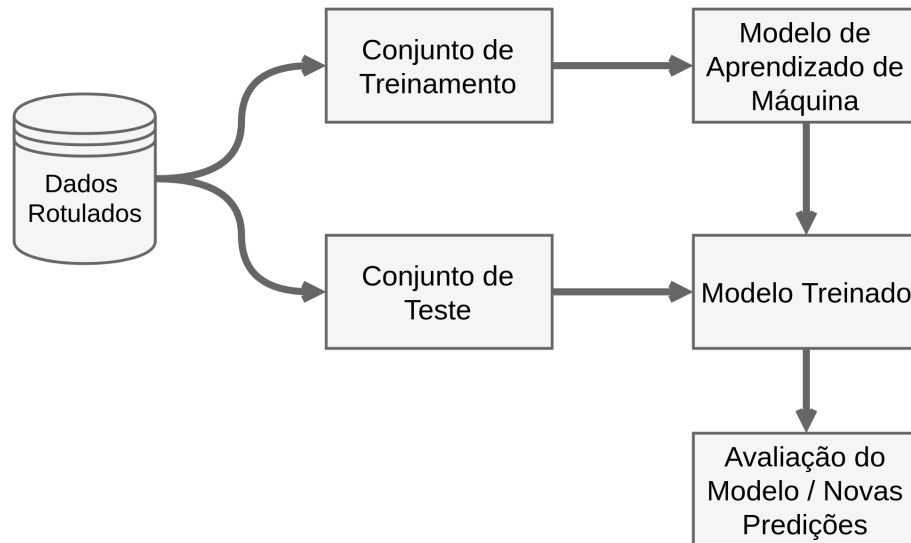


Figura 2: Pipeline do aprendizado de máquina supervisionado.

2.2 APRENDIZADO ATIVO

Considerando a grande quantidade de dados não rotulados disponíveis em relação à escassez de dados rotulados, bem como a necessidade de anotação dos dados por um especialista, torna-se essencial o desenvolvimento de abordagens de aprendizado mais robustas.

Aprendizado Ativo, sub-área do aprendizado de máquina supervisionado, tem como princípio abordagens para seleção de amostras mais informativas para o aprendizado do classificador, o qual auxilia ativamente no processo de aprendizado, selecionando e pré-rotulando tais amostras para posterior confirmação de rótulos por especialistas.

Estratégias de aprendizado ativo têm sido desenvolvidas, explorando diferentes critérios de seleção, de forma a obter acurácias de classificação elevadas mais rapidamente, com menos iterações de aprendizado e com o envolvimento mínimo do especialista no processo de anotação.

A Figura 3 ilustra o processo de aprendizado ativo tradicional. As principais técnicas de aprendizado ativo, as quais serão utilizadas como *baseline* nos experimentos, são apresentadas nas Subseções 2.2.1-2.2.6.

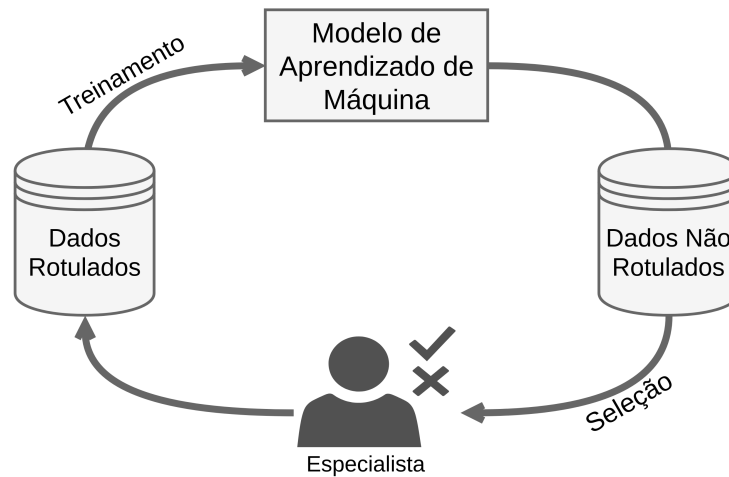


Figura 3: Processo de aprendizado ativo tradicional.

2.2.1 TÉCNICAS CLÁSSICAS BASEADAS EM MEDIDAS DE INCERTEZA

Existem técnicas clássicas que baseiam-se no critério de incerteza para selecionar amostras mais informativas a cada iteração do ciclo de aprendizado, dentre elas: *Least Confidence* (LC), *Margin Sampling* (MS), *Entropy* (EN) (SETTLES, 2012).

1. *Least Confidence* (LC) ou também conhecida como Menos Confiável é a mais simples dentre as clássicas, definida pela Equação 1.

$$LC(x) = 1 - P(y'|x) \quad (1)$$

em que x é uma amostra e y' é o rótulo mais provável atribuído pelo modelo. A estratégia prioriza as amostras com a maior incerteza da classe mais provável. Supondo um problema de classificação com as classes: *car*, *truck* e *bike*, ao classificar uma amostra A_1 obtém-se as seguintes probabilidades: *car* = 0.1, *truck* = 0.2, *bike* = 0.7 e ao classificar uma outra amostra A_2 obtém-se: *car* = 0.6, *truck* = 0.35, *bike* = 0.05.

Sendo assim, a classe mais provável para a amostra A_1 é a *bike*, com um valor de incerteza 0.3 ($LC(A_1) = 1 - 0.7$). No caso da amostra A_2 a classe mais provável é a *car* com valor de incerteza 0.4. Ou seja, considerando as amostras A_1 e A_2 a amostra A_2 possui o maior valor de incerteza, portanto seria a primeira a ser selecionada por essa estratégia. Uma desvantagem dessa estratégia é que ela só considera a informação da melhor previsão de um classificador, por exemplo.

2. *Margin Sampling* (MS) é definida pela Equação 2.

$$MS(x) = P(y''|x) - P(y'|x) \quad (2)$$

que avalia as duas maiores probabilidades de rótulo (y' e y'') de uma amostra (x) dada uma respectiva instância de um classificador, por exemplo. Prioriza as amostras com uma margem menor entre as duas classes mais prováveis. Considerando o mesmo exemplo mencionado anteriormente, com as classes: *car*, *truck* e *bike*, ao classificar uma amostra A_1 obtém-se as seguintes probabilidades: $car = 0.1$, $truck = 0.2$, $bike = 0.7$ e ao classificar uma outra amostra A_2 obtém-se: $car = 0.6$, $truck = 0.35$, $bike = 0.05$.

Dessa forma, a classe mais provável para a amostra A_1 é a *bike* seguida da segunda classe mais provável *truck*, sendo o valor da incerteza, segundo a MS, de -0.5 ($MS(A_1) = 0.2 - 0.7$). No caso da amostra A_2 a classe mais provável é a *car* seguida da segunda classe mais provável *truck*, com valor de incerteza -0.25 . Então, considerando as amostras A_1 e A_2 , a amostra A_2 possui o maior valor de incerteza, portanto seria a primeira a ser selecionada por essa estratégia. Ainda que essa estratégia não utilize somente a informação do rótulo do melhor palpite do classificador, para problemas com grande número de classes muita informação é ignorada.

3. *Entropy* (EN) ou Entropia é definida pela Equação 3.

$$EN(x) = - \sum_y P(y|x) \log P(y|x) \quad (3)$$

onde y denota todas as probabilidades de rótulos de uma amostra x . O método prioriza a amostra com maior entropia para seleção. A entropia, ao contrário das medidas anteriores, não favorece apenas as instâncias para as quais apenas um dos rótulos é altamente improvável. Um exemplo com as classes: *car*, *truck* e *bike*, ao classificar uma amostra A_1 obtém-se as seguintes probabilidades: $car = 0.1$, $truck = 0.2$, $bike = 0.7$ e ao classificar uma outra amostra A_2 obtém-se: $car = 0.6$, $truck = 0.35$, $bike = 0.05$.

Segundo a Equação 3, o valor de entropia para a amostra A_1 é 0.8, enquanto que o valor de entropia para a amostra A_2 é de 0.82. Portanto, considerando as duas amostras A_1 e A_2 , a amostra A_2 seria a selecionada por possuir maior valor de entropia. A entropia é apropriada se a função objetivo é minimizar a perda logarítmica que é uma medida de desempenho de um modelo de classificação. Já os outros dois métodos são mais apropriados se o objetivo for reduzir o erro de classificação, uma vez que priorizam instâncias que ajudariam o modelo a distinguir melhor classes específicas.

2.2.2 UNSUPERVISED CLUSTERING (UC)

A técnica de aprendizado ativo *Unsupervised Clustering* (UC) (ALAJLAN et al., 2014) realiza um pré-processamento, organizando os dados de maneira não supervisionada, por meio

de métodos de *clustering* para agrupar os dados. Após essa etapa, a técnica UC utiliza as amostras raízes dos *clusters* para compor o primeiro conjunto de amostras para anotação pelos especialistas e posteriormente o treinamento do classificador.

Nas demais iterações as amostras de cada *cluster* são selecionadas de maneira aleatória. O objetivo principal da técnica UC é explorar a diversidade dos dados, porém critérios de incerteza e representatividade dos dados não são explorados.

2.2.3 RANDOM BOUNDARY EDGES (RBE)

Na estratégia de aprendizado ativo *Random Boundary Edges* (RBE), inicialmente, os dados são pré-processados e organizados em *clusters* (SAITO et al., 2012). Em seguida, explora-se o critério de incerteza ao selecionar amostras de fronteiras entre os diferentes *clusters*. Esse critério de seleção reforça que amostras mais difíceis para classificação sejam escolhidas para etapa de treinamento do algoritmo. Outro ponto importante dessa técnica é que não é necessária a reorganização das amostras a cada iteração, dado que é proposta a redução e organização a priori, como um pré-processamento realizado uma única vez.

Além disso, a cada iteração não ocorre a classificação de todas as amostras do conjunto, apenas um pequeno conjunto é classificado e selecionado, tais pontos tornam-na adequada para tarefas que exigem um tempo de resposta rápido mesmo quando consideramos sua utilização em conjuntos de dados com grande quantidade de amostras. Esse método, apesar de explorar o critério de incerteza, não considera critérios de diversidade e representatividade para seleção de amostras mais informativas.

A mecânica do aprendizado com RBE, em sua primeira iteração, funciona de maneira semelhante à técnica de UC, ou seja, seleciona as amostras correspondentes às raízes dos *clusters*. Essas amostras são apresentadas ao especialista para anotação dos dados e em seguida para treinamento inicial do classificador. Nas iterações que seguem, outras amostras de fronteira são selecionadas de forma aleatória e apresentadas ao especialista que pode confirmar ou corrigir o pré-rótulo que as instâncias anteriores dos classificadores atribuíram. Esse processo continua até que o especialista esteja satisfeito com a eficácia do classificador.

2.2.4 DECREASING BOUNDARY EDGES (DBE)

A estratégia *Decreasing Boundary Edges* (DBE) (SAITO et al., 2013), na etapa inicial, assim como na RBE, organiza o conjunto em *clusters* seguido da redução dos dados. Ao agrupar o conjunto e selecionar amostras de fronteira, explora-se o critério de incerteza. Amostras de

fronteiras são organizadas com base na ordem decrescente de peso de suas arestas. Nesse caso, as maiores arestas são priorizadas de forma que classes distintas estejam presentes no conjunto a ser selecionado, ou seja, explora-se também o critério de diversidade.

O DBE não exige classificação e reorganização de todas as amostras não anotadas no conjunto de dados em cada iteração. Por esta razão, a estratégia é eficiente mesmo para grandes conjuntos de dados. Na primeira iteração de aprendizado, as amostras raízes de cada *cluster* são selecionadas. Estas amostras anotadas constituem o conjunto de treinamento da primeira instância do classificador. Durante as demais iterações de aprendizado, as amostras pré-organizadas com base na ordem decrescente das arestas a partir de cada *cluster* são selecionadas. O processo é iterativo e finaliza quando o especialista estiver satisfeito com o desempenho do classificador.

2.2.5 MINIMUM-SPANNING TREE BOUNDARY EDGES (MST-BE)

A estratégia de aprendizado ativo *Minimum Spanning Tree Boundary Edges* (MST-BE) (SAITO et al., 2014a) considera que os dados já foram pré-processados e organizados em *clusters* e, assim como no RBE, também explora o critério de incerteza ao selecionar amostras de fronteiras entre os diferentes *clusters*. O que diferencia a técnica MST-BE é a forma de organização e seleção das amostras de fronteira, que obtém um grafo completo formado pelas amostras de fronteira, em que cada aresta é ponderada por pesos que representam a distância entre as amostras. Em seguida, é calculada a *Minimum Spanning Tree* (MST), por meio do grafo completo das amostras de fronteira. As amostras (arestas) da MST são organizadas em ordem decrescente de pesos das arestas.

A técnica prioriza a seleção de amostras de fronteira com a menor distância (arestas mínimas) e classes distintas (maiores arestas da MST) do grafo completo. A intuição por trás dessa estratégia é que amostras “mais próximas” normalmente pertencem a uma mesma classe, porém quando isso não é verdade, pode-se atribuir alto nível de relevância e informação para essas amostras elevando a qualidade do classificador durante a etapa de treinamento.

A técnica MST-BE, em sua primeira iteração, seleciona as amostras raízes dos *clusters*. Essas amostras são apresentadas ao especialista para anotação dos dados e seguidas para treinamento inicial do classificador. Nas iterações que seguem, a instância atual do classificador recebe um par de amostras (arestas do grafo ordenadas de forma decrescente) e se essas amostras pertencerem a classes distintas elas são apresentadas ao especialista que confirma ou atribui o rótulo correto e segue o treinamento de uma nova instância do classificador. Caso o par de amostras seja classificado como pertencente a uma mesma classe, o algoritmo obtém

o próximo par da lista ordenada para avaliação. Apesar da estratégia MST-BE considerar o critério de incerteza, critérios de diversidade e representatividade poderiam também ser melhor explorados.

2.2.6 ROOT-DISTANCE BASED SAMPLING (RDS)

Na estratégia de aprendizado ativo *Root Distance-Based Sampling* (RDS) (SAITO et al., 2015), inicialmente, os dados são pré-processados e organizados em *clusters*. Em seguida, uma lista para cada *cluster* é criada, sendo as amostras de cada *cluster* armazenadas em suas respectivas listas, em ordem crescente de distância em relação à amostra raiz do seu respectivo *cluster*.

A técnica RDS explora o critério de diversidade ao selecionar a cada iteração amostras de grupos (listas) distintos. Além disso, é considerado o critério de incerteza, em que o classificador, ao auxiliar o processo de aprendizado, seleciona amostras mais difíceis (incertas) de serem classificadas. Nesse sentido, para cada lista, uma dada amostra é selecionada apenas se o rótulo, fornecido pela instância atual do classificador, for distinto do rótulo da amostra raiz do respectivo grupo (lista).

A técnica RDS leva em consideração os critérios de diversidade e incerteza. No entanto, esses critérios, entre outros, podem ser melhor explorados, levando em consideração as interações do especialista, em iterações anteriores, no processo de anotação. Portanto, melhorias nessas estratégias de seleção podem também ser melhor investigadas no presente projeto, de forma a melhorar o processo de aprendizado.

2.3 APRENDIZADO PROFUNDO

Aprendizado profundo, uma sub-área do aprendizado de máquina, trata de redes neurais artificiais complexas, isto é, com um grande número de camadas e conexões. Uma rede neural artificial é composta por um conjunto de unidades conectadas, também chamados de neurônios artificiais (análogo aos neurônios biológicos em um cérebro animal). Cada conexão (análoga a uma sinapse) entre neurônios artificiais pode transmitir um sinal entre eles. O neurônio artificial que recebe o sinal pode processá-lo e depois propagá-lo para os neurônios de camadas subsequentes conectados a ele. Entretanto, as redes neurais profundas atuais vão muito além da simplicidade reproduzida em uma rede *Perceptron* (ROSENBLATT, 1958), uma vez que é um conjunto de técnicas projetadas para permitir que modelos computacionais aprendam representações de dados em níveis de abstração semelhantes ao cérebro humano (LECUN

et al., 2015).

A Figura 4 ilustra a composição de uma rede neural simples e uma rede neural profunda que possui múltiplas camadas ocultas.

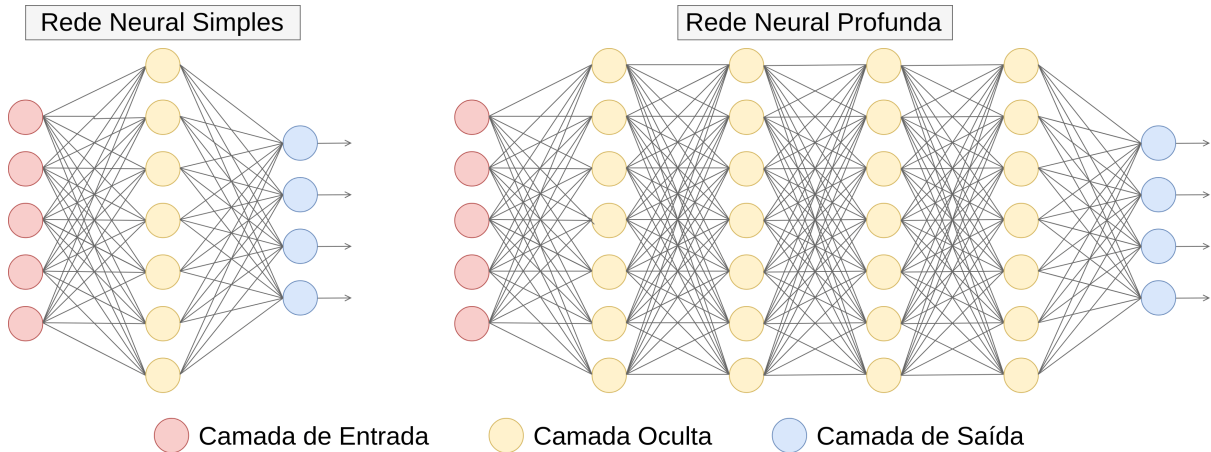


Figura 4: Representação de uma Rede Neural Simples (esquerda) e uma Rede Neural Profunda com múltiplas camadas ocultas (direita).

Em geral, o aprendizado profundo é capaz de resolver problemas de classificação de imagens melhor do que qualquer outro método paramétrico ou não paramétrico na literatura (LECUN et al., 2015). No entanto, existe um *trade-off* entre o custo computacional e a acurácia. Redes mais complexas e profundas necessitam de mais tempo na etapa de treinamento, exigindo um grande poder computacional, em contrapartida, de forma geral obtêm melhores resultados. Além disso, a disponibilidade de um grande conjunto de dados (por exemplo, mais de 5.000 amostras por classe) é outro ponto essencial para a obtenção de resultados positivos (GOOD-FELLOW et al., 2016).

2.3.1 CONVOLUTIONAL NEURAL NETWORK (CNN)

As *Convolutional Neural Networks* (CNNs) contemplam o estado da arte para a classificação de imagens utilizando aprendizagem profunda (LECUN et al., 2015) (RAWAT; WANG, 2017). Como em qualquer rede neural, as redes convolucionais são compostas por neurônios que recebem dados de entrada e executam a operação de produto escalar com os pesos, seguido de uma função de ativação. Uma das diferenças entre uma rede neural clássica e as CNNs é a redução no número de conexões entre os neurônios de cada camada, devido às operações de convolução e subamostragem (*pooling* ou *subsampling*). Portanto, apresenta um custo computacional reduzido quando comparada a uma arquitetura padrão de rede neural totalmente conectada.

A operação de convolução tem um papel importante na preservação da estrutura dos dados de entrada, ao contrário de uma rede neural totalmente conectada, na qual qualquer forma de entrada é sempre convertida em vetor de uma única dimensão. A preservação da estrutura de entrada ajuda a visualizar as saídas da camada intermediária nas CNNs e permite fazer interpretações e observações úteis em diferentes camadas da rede.

A Figura 5 ilustra um exemplo da operação de convolução aplicada a um conjunto de *pixels* de uma imagem digital. A operação de convolução, no contexto das CNNs e de maneira simplificada, pode ser denotada como a soma do produto de um determinado conjunto de *pixels* sobrepostos.

A operação de *max pooling*, representada na Figura 6, tem o papel de fornecer discreta invariância a translações nos dados de entrada e diminuir o número de parâmetros da rede, reduzindo o custo computacional. No entanto, a operação de *pooling*, apesar de apresentar bons resultados na prática em uma grande variedade de aplicações e arquiteturas de CNNs, ainda não se tem uma analogia de tal mecanismo presente no cérebro humano, o que leva alguns autores a criticarem tal técnica (SABOUR et al., 2017).

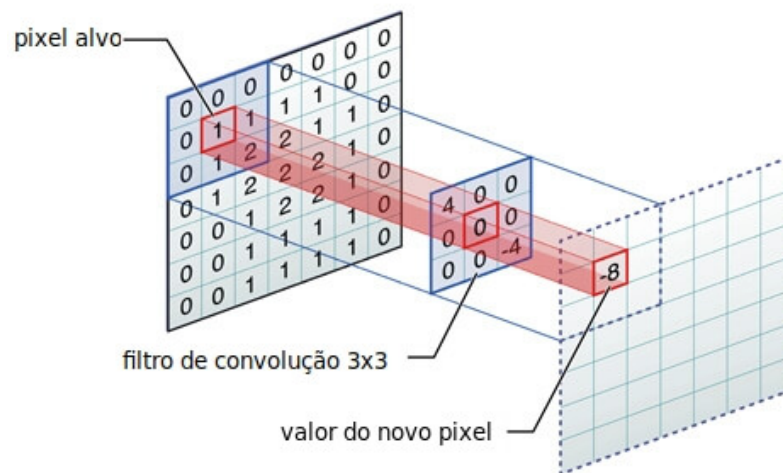


Figura 5: Representação da operação de convolução. O elemento central do filtro é colocado sobre o pixel alvo, o qual é substituído pela soma ponderada do mesmo e seus vizinhos.

A arquitetura LeNet-5 (LECUN et al., 1998), ilustrada na Figura 7, foi o primeiro caso bem sucedido na aplicação de CNNs para classificar imagens de caracteres manuscritos. É uma arquitetura simples que consiste em apenas duas camadas de convolução com *pooling* (subamostragem), seguida por duas camadas totalmente conectadas.

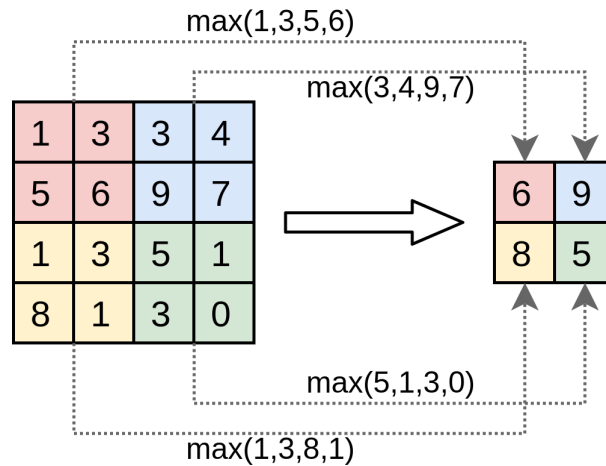


Figura 6: Representação da operação de *max pooling* com filtro de tamanho 2×2 e tamanho do passo 2.

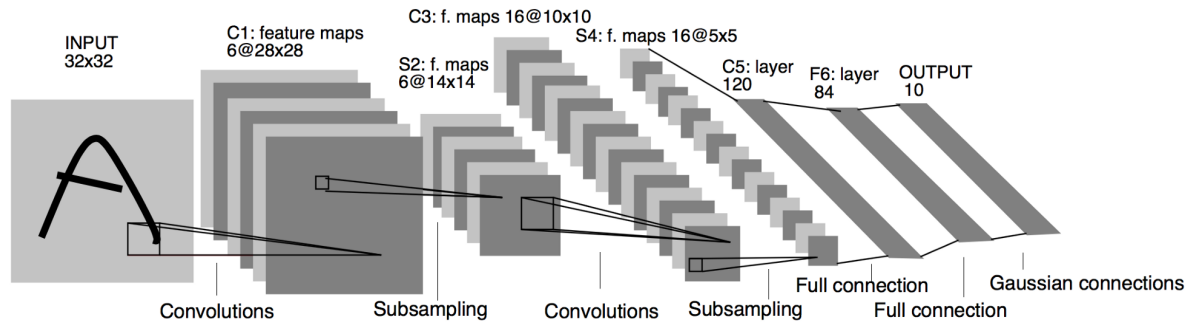


Figura 7: Arquitetura LeNet-5 composta por 2 camadas de convolução com subamostragem e 2 camadas totalmente conectadas.

Fonte: (LECUN et al., 1998)

2.3.1.1 ALEXNET

O desenvolvimento da arquitetura AlexNet (KRIZHEVSKY et al., 2012), foi um marco importante para a era do aprendizado profundo na área de visão computacional, pois foi pioneira nesse sentido, e precursora das demais arquiteturas de CNNs.

A arquitetura AlexNet (KRIZHEVSKY et al., 2012), além de ser a vencedora da competição ILSVRC em 2012, reduzindo a taxa de erro em aproximadamente 10% comparado ao ano anterior, também aplicou novos conceitos para as redes neurais convolucionais. Dentre eles, a técnica de regularização denominada de *Dropout* (SRIVASTAVA et al., 2014), a qual se mostrou eficiente na redução do *overfitting*, destacada na Figura 8.

A técnica de *Dropout* é responsável por diminuir a capacidade de memorização e aumentar o poder de generalização de uma rede com muitos parâmetros, como é o caso da Alex-

Net. Simplificadamente, o *Dropout* é responsável por zerar a saída de um conjunto de neurônios aleatório. A quantidade desse grupo de neurônios é definida por um parâmetro de entrada, normalmente uma probabilidade.

Outra importante contribuição foi o uso da função de ativação *Rectified Linear Unit* (ReLU), dada por $f(x) = \max(0, x)$, que oferece redução no tempo de treinamento da rede em comparação com as tradicionais funções *tanh* $f(x) = -1 + 2/(1 + e^{-2x})$ e *sigmoid* $f(x) = (1 + e^{-x})^{-1}$. A ReLU também reduz a chance de ocorrer o desaparecimento do gradiente (BENGIO et al., 1994) durante o treinamento da rede com *backpropagation* (RUMELHART et al., 1986). A Figura 9 mostra a representação gráfica dessas funções.

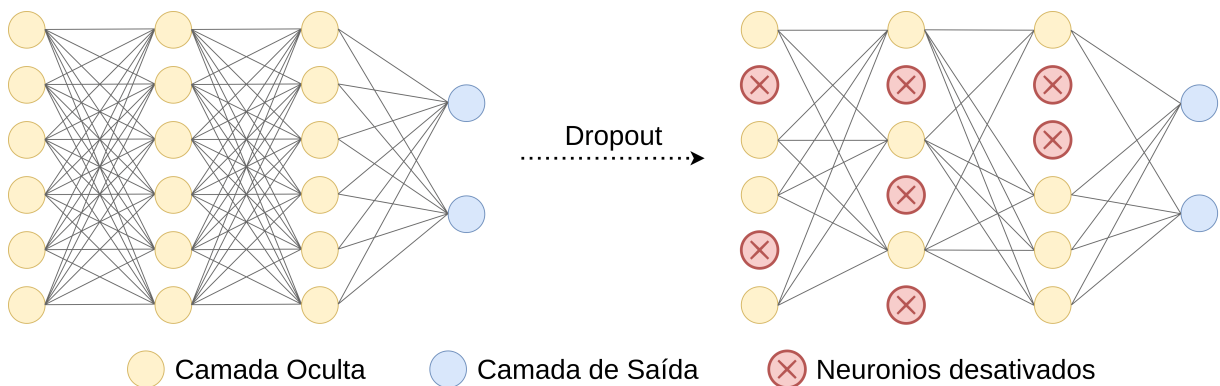


Figura 8: Ilustração da técnica de regularização *Dropout*.

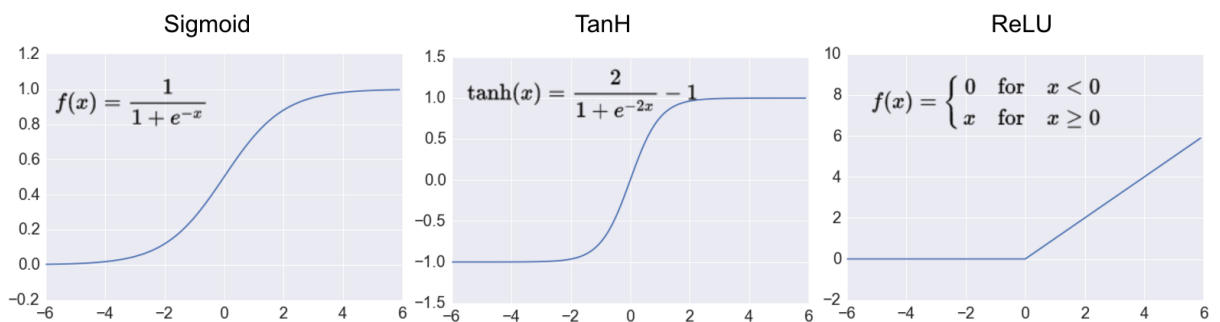


Figura 9: Principais funções de ativação para as redes neurais.

A rede neural na arquitetura AlexNet apresenta 60 milhões de parâmetros e 650 mil neurônios, composta por cinco camadas de convolução, algumas das quais são seguidas por camadas de *pooling*, três camadas totalmente conectadas e por fim a função *softmax* que retorna um vetor de 1000 posições na saída. A Figura 10 ilustra a arquitetura da AlexNet.

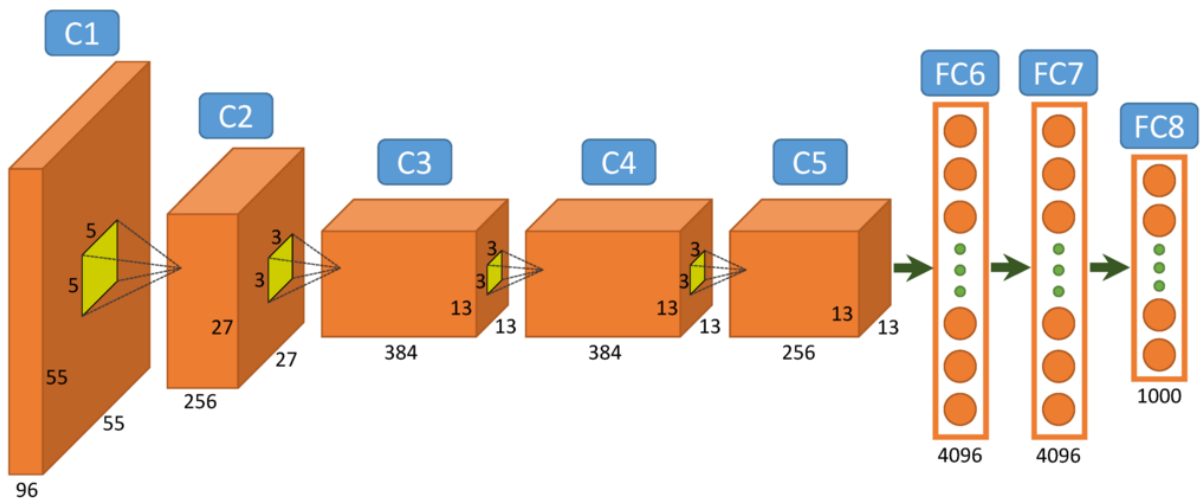


Figura 10: Arquitetura AlexNet, composta por 5 camadas de convolução e 3 camadas totalmente conectadas.

Fonte: Disponível em: <https://www.saagie.com/blog/object-detection-part1>.

2.3.1.2 GOOGLNET / INCEPTION

Inicialmente, conhecido como *GoogLeNet* (SZEGEDY et al., 2015), um modelo de rede introduzido e vencedor do ILSVRC em 2014. Esse modelo mostra-se eficiente, ao propor uma redução no número de parâmetros (pesos) da rede, quando comparado às redes do estado da arte anteriores. O sucesso da arquitetura está relacionada ao inovador uso dos blocos chamados de *Inception*, nos quais as operações de convolução e *pooling* ocorrem em paralelo, e em seguida as saídas são concatenadas. A Figura 11 ilustra as camadas e blocos *Inception* do modelo *GoogLeNet*.

Posteriormente, em Ioffe e Szegedy (2015), foi introduzido o conceito de *Batch Normalization* (BN) para a arquitetura e em Szegedy et al. (2016) foram apresentadas as variantes *Inception-v2* e *Inception-v3*. Na *Inception-v2* as operações de convolução com filtro de tamanho 7×7 foram substituídas por 3 operações de convolução com filtro de tamanho 3×3 . Na *Inception-v3* passaram a utilizar *Batch Normalization* também na saídas dos classificadores auxiliares da rede.

A *Inception-v4* (SZEGEDY et al., 2017) combinou o conceito de conexões em paralelo (blocos de *Inception*) com o conceito de conexões residuais (vide Subseção 2.3.1.3). Foi evidenciado que essa combinação, além de melhorar significativamente a velocidade de treinamento da rede, quando comparada a uma arquitetura *Inception* tradicional (sem o uso de conexões residuais), também ocorre uma melhoria discreta da acurácia.

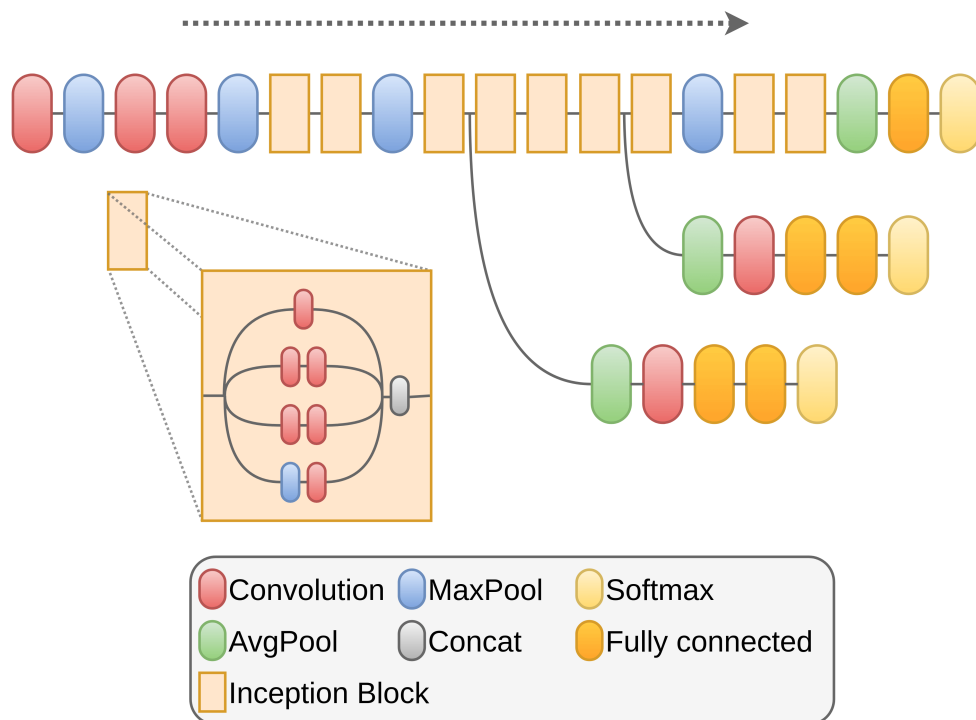


Figura 11: Arquitetura *GoogleNet* ou *Inception-v1* composta por 22 camadas, vencedora da tarefa de classificação na competição ILSVRC de 2014 com uma taxa de erro de 6,7%.

2.3.1.3 RESIDUAL NETWORK (RESNET)

A arquitetura *Residual Network* - ResNet (HE et al., 2016a) explora o conceito e potencial da profundidade das redes neurais, ou seja, redes com um grande número de camadas. A ResNet introduziu uma arquitetura com 152 camadas que a conduziu para o novo estado da arte do ano de sua publicação. A ResNet foi vencedora, com uma margem considerável, nas tarefas de classificação, localização e detecção de objetos na ILSVRC de 2015.

Os autores iniciaram com a premissa, e comprovaram por meio de experimento ilustrado na Figura 12, de que para criar melhores arquiteturas não é tão simples quanto aumentar o número de camadas.

A arquitetura ResNet traz o conceito de conexão residual em que, dada uma entrada x , a mesma é somada às saídas das duas camadas subsequentes, conforme ilustra a Figura 13. Com as conexões residuais foi possível criar modelos com elevado número de camadas (crescimento em profundidade), melhorando a taxa de erro durante o treinamento da rede.

Em He et al. (2016b) novas análises e experimentos com a ResNet foram realizados e concluíram que os sinais dos blocos residuais podem ser propagados diretamente para qualquer outro bloco, utilizando mapas de identidade, proporcionando melhorias durante o treinamento e na generalização quando comparada com a versão original.

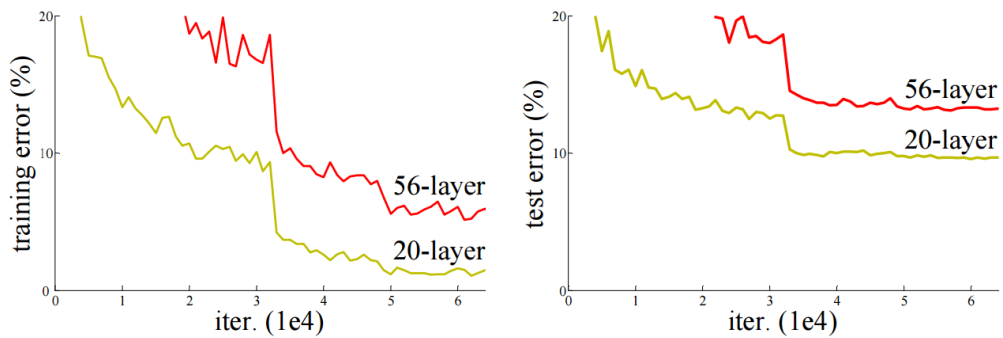


Figura 12: Taxa de erro durante o treinamento (esquerda) e taxa de erro durante o teste (direita).

Fonte: He et al. (2016a)

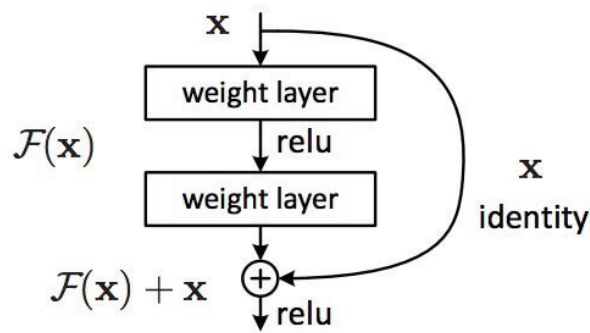


Figura 13: Bloco de conexão residual da arquitetura ResNet.

Fonte: He et al. (2016a)

2.3.1.4 ALÉM DAS CNNs

Embora as redes neurais convolucionais tenham sido o destaque para resolver problemas de classificação e detecção de objetos em imagens e sejam boas em detectar características locais, a técnica é menos eficaz na detecção de relações espaciais entre diferentes objetos de uma imagem por exemplo. Parte desse problema é causado pela detecção invariante de características. As CNNs processam a probabilidade de objetos em uma imagem sem processar as propriedades desses objetos (por exemplo, ângulo ou tamanho).

Em Sabour et al. (2017) foi proposto um novo modelo capaz de superar algumas das limitações presentes nas arquiteturas de redes neurais convolucionais. Dentre essas limitações, a principal foi que, diferentemente das redes neurais convolucionais, as quais normalmente utilizam um conjunto de operações de *pooling* e com isso perdem informações de localização e inclinação de um dado objeto em uma imagem, as chamadas *Capsule Networks* são capazes de preservar essas propriedades. Uma rede de cápsula é composta por várias cápsulas, onde a

cápsula pode ser qualquer função que tenta prever a presença e os parâmetros de criação de um dado objeto em uma imagem.

Para as tarefas de detecção de objetos e segmentação de imagens essa propriedade de equivariância das *Capsule Networks* as tornam muito promissoras. Foi reportado desempenho de estado da arte na tarefa de classificação de dígitos manuscritos no conjunto de dados MNIST (LECUN et al., 1998), também foram demonstrados resultados consideravelmente melhores do que as redes neurais convolucionais, quando aplicados a dígitos altamente sobrepostos.

2.3.2 GENERATIVE ADVERSARIAL NETWORK (GAN)

As *Generative Adversarial Networks* (GANs), propostas por (GOODFELLOW et al., 2014), são uma maneira de aprender representações dos dados sem que esses estejam rotulados. As GANs evoluíram para uma das mais bem-sucedidas técnicas de aprendizado não supervisionadas para geração de imagens realistas (ZHU et al., 2017).

A Figura 14 ilustra o modelo geral das GANs. Basicamente, duas redes neurais são treinadas, de forma a competir uma com a outra, sendo uma delas denominada de Gerador (G) e a outra de Discriminador (D).

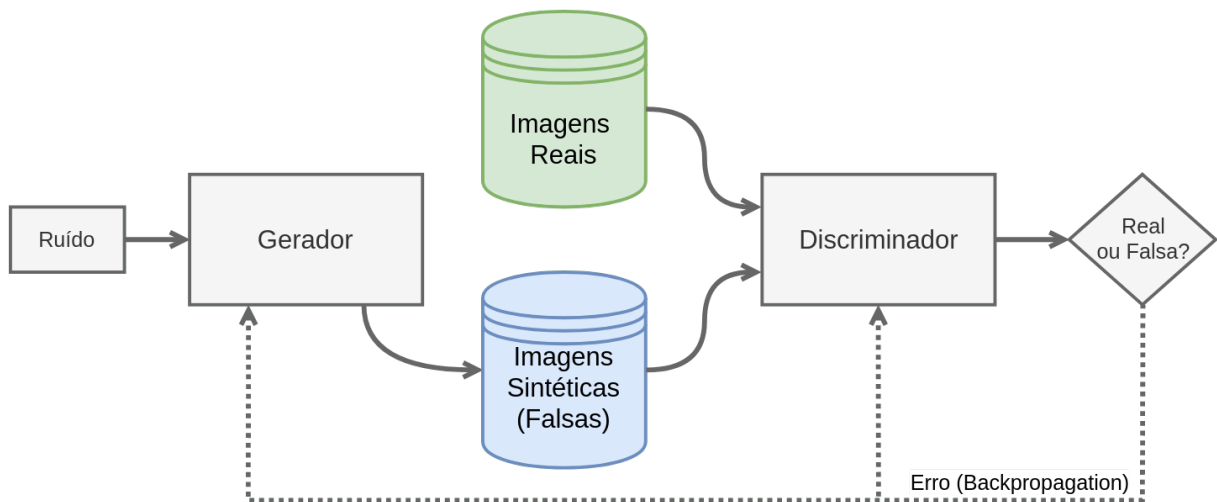


Figura 14: Modelo das Generative Adversarial Networks (GANs).

O Gerador é responsável por gerar novas amostras que serão dados de entrada para o Discriminador. Dado que não possui acesso aos dados da base real, a única forma de aprendizagem é por meio de sua interação com o Discriminador. A entrada inicial para treinamento do Gerador, normalmente, é ruído na forma de uma distribuição gaussiana. O objetivo do Gerador é enganar o Discriminador, de maneira que o mesmo não consiga detectar que trata-se de um dado sintético.

O Discriminador atua como um classificador binário, o qual é treinado para distinguir entre amostras da base de dados (também denominadas de amostras reais) e amostras sintéticas (também denominadas de amostras falsas).

O treinamento das redes envolve encontrar os parâmetros que maximizam a acurácia do discriminador e ao mesmo tempo maximizam a capacidade do Gerador em enganar o Discriminador, ao tentar fazer uma imagem falsa se passar por uma real. Durante cada etapa do treinamento, o erro do Discriminador também é propagado para o Gerador, fazendo com que o Gerador seja capaz de produzir imagens o mais fidedignas possível, quando comparadas com as imagens encontradas na base original.

A proposta original da GAN (GOODFELLOW et al., 2014) pode ser definida pela Equação 4, onde o componente Discriminador $D : X \rightarrow [0, 1]$ e o Gerador $G : Z \rightarrow X$, Z é um espaço latente e X o espaço das imagens reais do conjunto. \mathbb{P}_z é uma distribuição do espaço Z e \mathbb{P}_r é uma distribuição do espaço X .

$$\min_G \max_D \ell = \mathbb{E}_{\mathbf{x} \sim \mathbb{P}_r} [\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z} \sim \mathbb{P}_z} [\log(1 - D(G(\mathbf{z})))] \quad (4)$$

Um dos grandes problemas na proposta original da GAN (GOODFELLOW et al., 2014) está relacionada à dificuldade no treinamento, no qual o modelo pode nunca aprender algo ou convergir para um mínimo local. Outro problema está relacionado ao modo de colapso, em que o Gerador gera saídas idênticas infinitamente.

Os modelos tradicionais de GANs, se treinados o suficiente, sempre acabam nos equilíbrios locais de Nash (DASKALAKIS et al., 2009) e normalmente estão associados ao modo de colapso. Em Unterthiner et al. (2017) foi proposta a *Coulomb GAN*, e foi demonstrado que o treinamento de uma GAN converge e possui apenas um equilíbrio de Nash, sendo esse o ótimo.

Apesar da principal aplicação das GANs consistir na síntese de imagens (DENTON et al., 2015; REED et al., 2016; ZHU et al., 2017), elas também podem ser utilizadas para outros propósitos como, por exemplo, classificação ou regressão por meio da extração de atributos do Discriminador após treinamento da rede (RADFORD et al., 2015; SALIMANS et al., 2016).

2.3.2.1 DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORK (DCGAN)

No contexto de síntese de imagens a proposta da GAN original (GOODFELLOW et al., 2014) trouxe resultados razoavelmente bons em um conjunto de dados simples (MNIST).

No entanto, quando utilizada em conjuntos de dados mais complexos, os resultados na maioria são insatisfatórios, gerando imagens irreconhecíveis. Boa parte dessa deficiência da proposta original da GAN se deve ao fato de que ambos, Gerador e Discriminador, usam a arquitetura perceptron de multicamadas.

As CNNs são amplamente utilizadas para aprendizado supervisionado em problemas de visão computacional. No trabalho de Radford et al. (2015) a configuração da GAN foi combinada com sucesso com uma CNN, originando a *Deep Convolution Generative Adversarial Network* (DCGAN).

A DCGAN, em conjunto com os avanços na área de visão computacional, foi capaz de melhorar a estabilidade da GAN por meio do uso de Redes Neurais Convolucionais, técnica de regularização *Batch Normalization* e também, com o intuito de não perder as informações espaciais dos objetos nas imagens, houve a substituição das camadas de *pooling* por *strides* (tamanho do passo) 2×2 nas operações de convolução. Além disso, demonstrou que os parâmetros aprendidos pelo Discriminador são ótimos candidatos para extração de características. A Figura 15 ilustra a arquitetura da DCGAN.

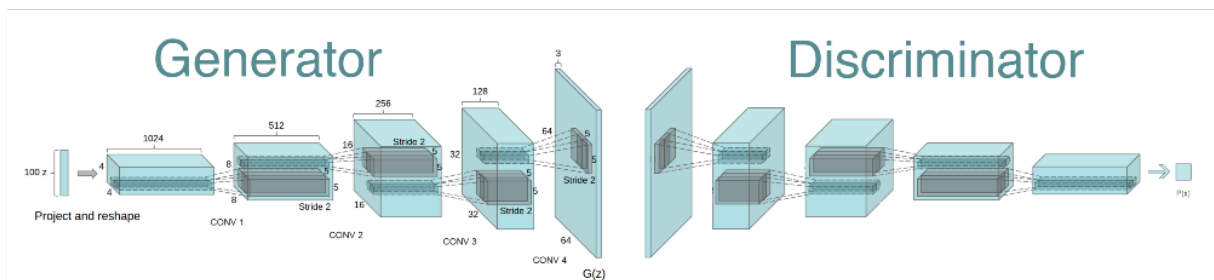


Figura 15: Arquitetura da DCGAN.

Fonte: Radford et al. (2015)

2.3.2.2 BOUNDARY EQUILIBRIUM GENERATIVE ADVERSARIAL NETWORK (BEGAN)

Em Berthelot et al. (2017) é proposta a *Boundary Equilibrium Generative Adversarial Network* (BEGAN) que utiliza um modelo de *autoencoder* (GOODFELLOW et al., 2016) para o discriminador de uma GAN. Diferentemente da GAN original, que usa o discriminador como um classificador binário, a arquitetura de *autoencoder* extrai as características das imagens através de um *encoder* e as reconstrói novamente através de um *decoder*.

A BEGAN usa um critério diferente da maioria das variantes de GANs propostas na literatura, que baseia-se em uma proposição de que há uma correspondência entre a distribuição

do erro e a distribuição dos dados de imagens reais e imagens geradas, ou seja, a BEGAN tenta encontrar a distribuição do erro do *autoencoder* (Discriminador) ao invés de tentar encontrar a distribuição dos dados diretamente. Para função custo ou erro da BEGAN, é utilizada uma variação da distância de *Wasserstein* (GULRAJANI et al., 2017) que também age como um medida de convergência.

Sucintamente, a BEGAN possibilita uma arquitetura simples, porém robusta, com treinamento e convergência relativamente rápidos. Um conceito de equilíbrio que tenta balancear as redes geradora e discriminadora durante o treinamento. Um mecanismo de controle entre a qualidade visual ou diversidade das imagens geradas.

2.3.2.3 MÉTRICAS DAS GANs

Um dos desafios nos cenários das GANs está na escassez de métricas que avaliam a relação da qualidade visual das imagens geradas pelas GANs com o desempenho das mesmas em uma tarefa específica. Ainda não existe um consenso e uma métrica de qualidade amplamente utilizada nesse contexto é uma área de pesquisa ainda em ascensão.

No entanto, a métrica mais popular no âmbito das GANs é denominada de *Inception Score* (IS) (SALIMANS et al., 2016). Essa métrica utiliza a arquitetura *Inception* (vide Subseção 2.3.1.2) pré treinada no conjunto de dados da *ImageNet* e calcula a Equação 5.

$$IS(\mathbb{P}_g) = e^{\mathbb{E}_{\mathbf{x} \sim \mathbb{P}_g} [KL(p_M(y|\mathbf{x}) || p_M(y))]} \quad (5)$$

onde $p_M(y|\mathbf{x})$ denota a distribuição de rótulo de \mathbf{x} .

Um valor elevado para essa métrica ocorre quando a rede *Inception* classifica uma imagem sintética com uma margem alta de confiança. Isso sugere que o modelo foi capaz de gerar imagens com alta qualidade e diversidade. Um dos pontos fracos dessa métrica é que ela avalia a qualidade das imagens geradas isoladamente, ou seja, não propõe nenhuma avaliação de similaridade entre as imagens geradas e as imagens reais.

Outra métrica importante é a métrica *Fréchet Inception Distance* (FID) (HEUSEL et al., 2017). Assim como na *Inception Score*, utiliza a rede *Inception* pré treinada no conjunto de dados da *ImageNet*, porém ao invés de utilizar essa rede para classificar as amostras, a FID usa para extrair descritores das imagens. A partir dos descritores a FID constrói uma distribuição gaussiana multivariada, definida pela Equação 6.

$$FID(x, g) = \|\mu_x - \mu_g\| + \text{Tr}(\Sigma_x + \Sigma_g - 2(\Sigma_x \Sigma_g)^{\frac{1}{2}}) \quad (6)$$

onde μ denota a média e Σ a covariância da distribuição.

A FID é mais tolerante a ruídos e captura melhor a diversidade das imagens quando comparada com a IS. Valores mais baixos da FID significam melhor qualidade e diversidade das imagens geradas.

2.4 TRANSFER LEARNING

Transfer Learning é um tópico popular em aprendizado de máquina e mineração de dados. Também conhecida como adaptação de domínio quando aplicada na área de processamento de linguagem natural. Normalmente, utiliza dois domínios, o de origem e o de alvo. O domínio de origem detém grande quantidade de dados de treinamento rotulados, enquanto o domínio alvo tem pouco ou nenhum dado de treinamento. O objetivo é usar os dados rotulados no domínio origem para ajudar no aprendizado no domínio alvo (TAYLOR; STONE, 2009; PAN et al., 2010).

Existem muitos tipos de conhecimento que podem ser transferidos do domínio origem para o domínio alvo. Por exemplo, uma maneira típica de trabalhar com a classificação de imagens é treinar e testar um modelo de aprendizagem no mesmo domínio de dados. No entanto, isso nem sempre é viável, especialmente quando se trabalha com redes neurais profundas que exigem uma grande quantidade de dados rotulados para a etapa de treinamento da rede. Nesse contexto, a técnica de *Transfer Learning* pode ser aplicada para resolver esse problema. Analogamente, pode-se considerá-la como reutilização de conhecimento prévio.

Ainda no âmbito das redes neurais, também é conhecido que as camadas iniciais de uma rede neural podem ser utilizadas como descritores mais simples (por exemplo, detectores de borda), enquanto as camadas mais profundas aumentam a abstração na detecção de padrões, concatenando os descritores das camadas anteriores, e com isso são capazes de identificar formas mais complexas como, por exemplo um rosto ou uma roda de carro (YOSINSKI et al., 2014). A Figura 16 ilustra o conceito aplicado à tarefa de classificação.

Existem algumas estratégias para usar as técnicas de *Transfer Learning*. Em Hinton et al. (2015) e (SHIN et al., 2016), essas estratégias foram avaliadas com redes neurais convolucionais, evidenciando poderosos descritores genéricos de imagem extraídos desse processo. Além disso, eles apontaram que esses descritores devem ser os principais candidatos em uma tarefa de classificação de imagem.

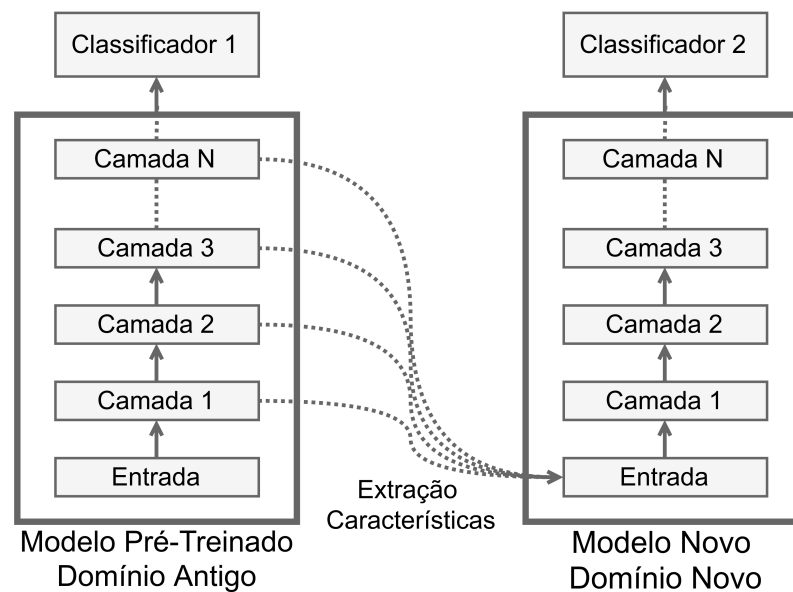


Figura 16: *Transfer Learning* aplicado à tarefa de classificação no contexto de redes neurais.

2.5 ABORDAGENS DE APRENDIZADO ATIVO E PROFUNDO

As abordagens de aprendizado ativo e profundo constituem um importante pilar do aprendizado de máquina. No entanto, existe uma carência da consolidação e utilização de tais técnicas em conjunto na literatura e embora as CNNs tenham obtido resultados impressionantes em diversas aplicações (SZEGEDY et al., 2016; HE et al., 2016a; HU et al., 2017), a utilização de estratégias de aprendizado ativo para esses modelos podem ser melhor exploradas.

Em Wang et al. (2017), é proposto um novo *framework* de aprendizado ativo, incorporando redes neurais convolucionais no aprendizado ativo. No *framework* proposto, a representação das características e o classificador podem ser alterados simultaneamente, a partir das amostras mais informativas anotadas e selecionadas pela estratégia de aprendizado ativo, durante o processo de aprendizado da rede. Além disso, é proposta uma nova estratégia de seleção de amostras, de forma a melhorar o desempenho da classificação com menos anotações manuais. Para tanto, a estratégia de seleção proposta, diferentemente das estratégias tradicionais, além de obter um pequeno conjunto composto pelas amostras mais incertas, também é obtido um grande conjunto constituído por amostras, as quais o classificador é mais confiante. O primeiro conjunto minoritário é exibido e anotado pelos especialistas. Para o segundo conjunto majoritário são atribuídos pseudo-rótulos fornecidos pela rede. Apesar dos resultados apresentados, como o conjunto majoritário recebe pseudo-rótulos fornecidos pela rede em um determinado estágio do aprendizado, erros podem ser propagados e considerados no treina-

mento. Além disso, a estratégia de aprendizado ativo proposta pelos autores considera todas as amostras do conjunto de dados a cada iteração do aprendizado. Sendo assim, o presente projeto tem como objetivo analisar estratégias de aprendizado ativo mais adequadas.

Em Gal et al. (2017), os autores propõem o *Deep Bayesian Active Learning*, combinando avanços recentes do aprendizado profundo Bayesiano para o *framework* de aprendizado ativo proposto, considerando imagens de alta dimensionalidade. Para o aprendizado ativo, são propostas três funções de aquisição para seleção das amostras. Os resultados demonstram que os modelos de redes neurais convolucionais Bayesianos, ao possibilitarem a propagação de incerteza durante o aprendizado, por meio das funções de aquisição, atingem e convergem para acurácias elevadas mais rapidamente, em relação aos modelos determinísticos de CNN. Foram também apresentadas as comparações entre a proposta de aprendizado ativo por meio de diferentes estratégias de aquisição e técnicas de aprendizado semi-supervisionado. As técnicas propostas foram avaliadas no conjunto de dados MNIST (LECUN et al., 1998) e em um conjunto para diagnóstico de câncer de pele (melanoma). Apesar dos resultados apresentados, os autores mencionam o custo de treinamento elevado para execução das funções de aquisição propostas. Por exemplo, para o conjunto melanoma, o tempo é de cerca de 20 horas para cada experimento. Os mesmos mencionam que tal tempo pode ser reduzido, caso o sistema não seja reiniciado. No entanto, nesse caso pode-se atingir um ótimo local. Sendo assim, tal problema deve ser abordado em trabalhos futuros.

Em Shen et al. (2017) é utilizada uma arquitetura que combina CNNs com *Long Short Term Memory* (LSTM) para a tarefa de reconhecimento de entidade mencionada (REM), aplicada em processamento de linguagem natural. Outros trabalhos recentes também adotaram técnicas que utilizam aprendizado profundo e aprendizado ativo aplicados a diferentes contextos, tais como: segmentação de imagens médicas (YANG et al., 2017; CHOWDHURY et al., 2017), geração de diálogos (ASGHAR et al., 2017) e navegação autônoma (HUSSEIN et al., 2016).

No entanto, geralmente, esses trabalhos envolvem a utilização de estratégias de aprendizado ativo, em que é necessário realizar a classificação, organização e seleção das amostras mais informativas, a cada iteração do aprendizado, considerando todo o conjunto de dados. Tal processo pode tornar-se bastante custoso e inviável de ser realizado. Portanto, explorar estratégias mais robustas e eficientes é fundamental.

3 DESCRIÇÃO DOS CONJUNTOS DE DADOS

Para avaliação e comparação das técnicas desenvolvidas nesse trabalho são utilizados três conjuntos de dados: *Canadian Institute For Advanced Research* (CIFAR), folhas de plantas e enteroparasitos, respectivamente.

CIFAR-10 (KRIZHEVSKY; HINTON, 2009) é um *dataset* clássico e muito utilizado na literatura para *benchmark* de diversos modelos da área de visão computacional, composto por 60.000 imagens com resolução 32×32 divididas em 10 classes de maneira balanceada. Ele está dividido em dois conjuntos, um para treinamento com 50.000 imagens e outro para testes composto de 10.000 imagens.

A Figura 17 apresenta exemplos de imagens e respectivas classes do conjunto de dados CIFAR-10. As classes presentes no *dataset* são mutualmente exclusivas, ou seja, não há sobreposição entre a classe caminhão (*truck*) e a classe automóveis (*automobile*).

O conjunto de folhas de plantas, LEA-53, é um subconjunto do Pl@ntLeaves II (GOËAU et al., 2012) para classificação do fenótipo de folhas. Este subconjunto consiste em 3.655 imagens de folhas segmentadas e com fundo branco divididas em 53 classes (espécies). Esse subconjunto foi escolhido por meio de um limiar definido empiricamente, em que as espécies presentes no conjunto original com número inferior a 30 amostras foram removidas.

Dessa forma, é possível conduzir experimentos em uma base minimamente balanceada, com mais classes e com mais amostras em cada classe, uma vez que as imagens foram divididas entre treinamento e teste na proporção 80% e 20%, respectivamente. As Figuras 18 e 19 ilustram, respectivamente, alguns exemplos de imagens e a distribuição das imagens em cada classe do subconjunto utilizado.

O conjunto de imagens de parasitos e impurezas intestinais (SUZUKI et al., 2013) encontra-se dividido em 3 grupos (ovos, larvas e protozoários). Sendo assim, é possível considerar diferentes subconjuntos de imagens. A Figura 20 ilustra exemplos de imagens presentes no conjunto de parasitos e impurezas intestinais. As quantidades de classes e imagens em cada um dos subconjuntos utilizados nos experimentos estão definidos na Tabela 1.

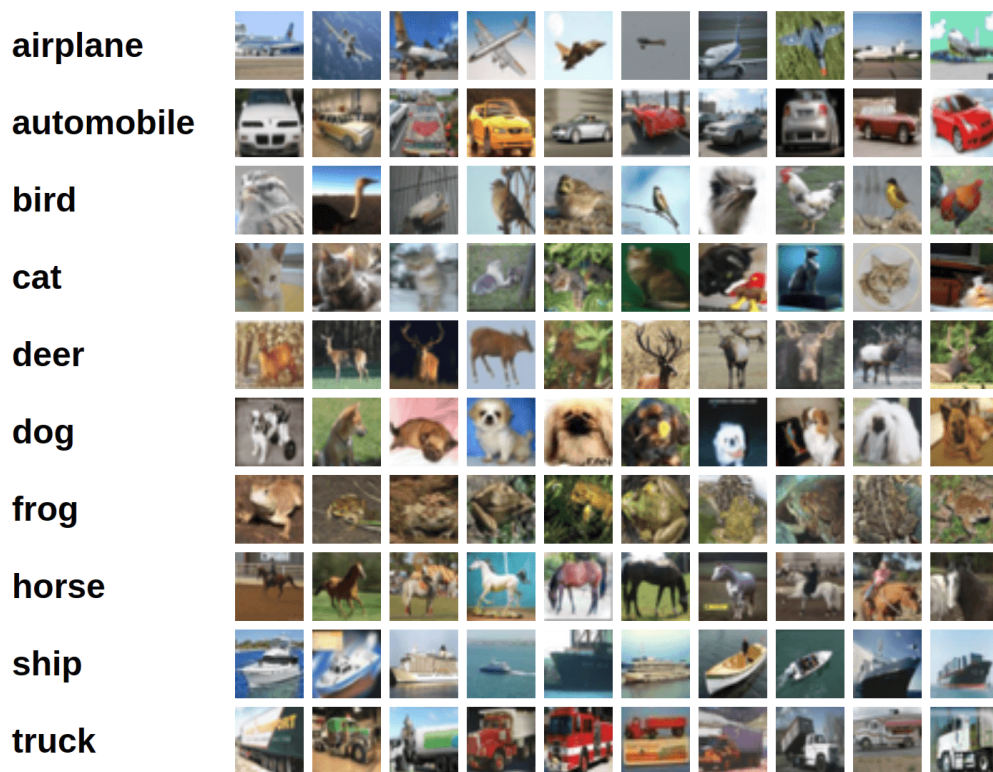


Figura 17: Exemplos de imagens e respectivas classes do conjunto de dados CIFAR-10.

Para o grupo de ovos foram considerados três subconjuntos (EGG-8, EGG-9 e EGG-16). EGG-8 corresponde ao conjunto de (8) espécies de enteroparasitos do grupo de ovos, sem a presença de classes de impurezas; EGG-9 compreende o conjunto de enteroparasitos EGG-8, juntamente com uma única classe de impurezas similares às espécies de enteroparasitos do grupo de ovos; EGG-16 refere-se ao conjunto de enteroparasitos do grupo de ovos EGG-8, juntamente com 8 classes de impurezas similares às (8) espécies de enteroparasitos do grupo de ovos. Referente ao grupo de larvas, o conjunto LAR-2 agrega uma classe de enteroparasitos e uma classe de impurezas similares à espécie de enteroparasitos desse grupo.

De forma análoga, para o grupo de protozoários, tem-se os subconjuntos PRO-6, PRO-7 e PRO-12. E, por fim, foram considerados três subconjuntos (PAR-15, PAR-16 e PAR-30), os quais abrangem amostras de todos os grupos (ovos, larvas e protozoários). PAR-15 apresenta amostras das 15 espécies de enteroparasitos, sem amostras referentes às impurezas; PAR-16 contém amostras do subconjunto PAR-15, juntamente com amostras referentes às impurezas constituindo uma única classe; e PAR-30 com amostras do subconjunto PAR-15, juntamente com amostras referentes às impurezas constituindo 15 classes diferentes). A Tabela 2 apresenta as quantidades de imagens em cada uma das classes consideradas nos (10) subconjuntos de dados utilizados. O conjunto de enteroparasitos é desafiador pois além da grande quantidade de amostras.



Figura 18: Exemplos de imagens do subconjunto de dados PI@ntLeaves II.

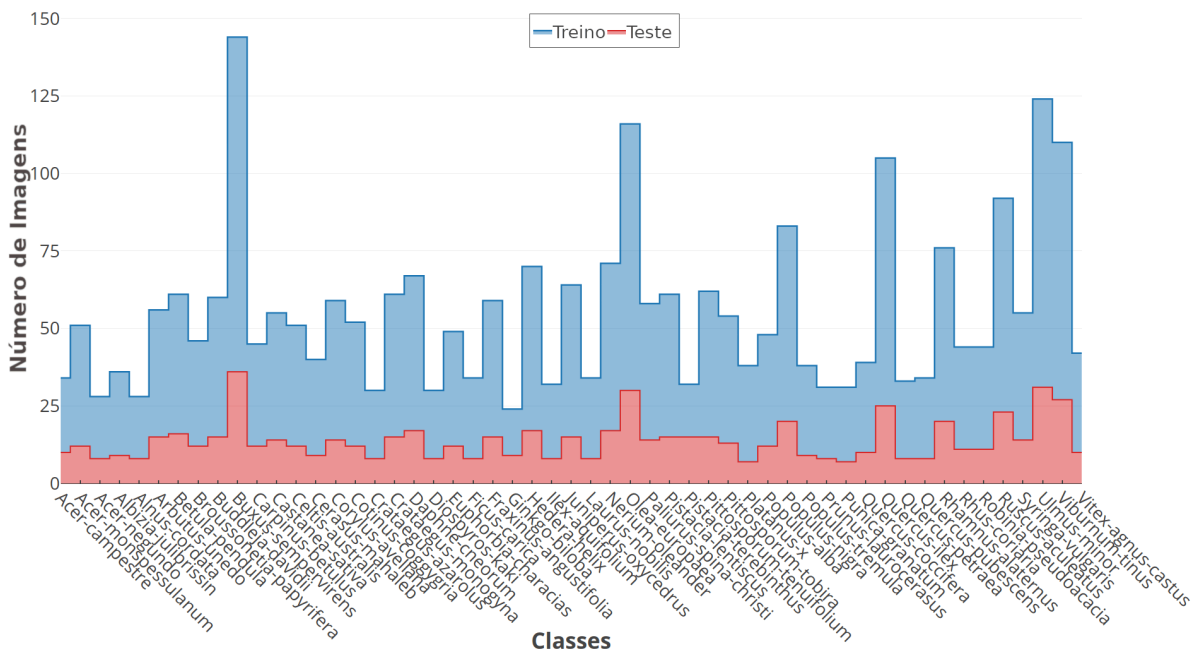


Figura 19: Distribuição das imagens em cada classe do subconjunto de dados PI@ntLeaves II.

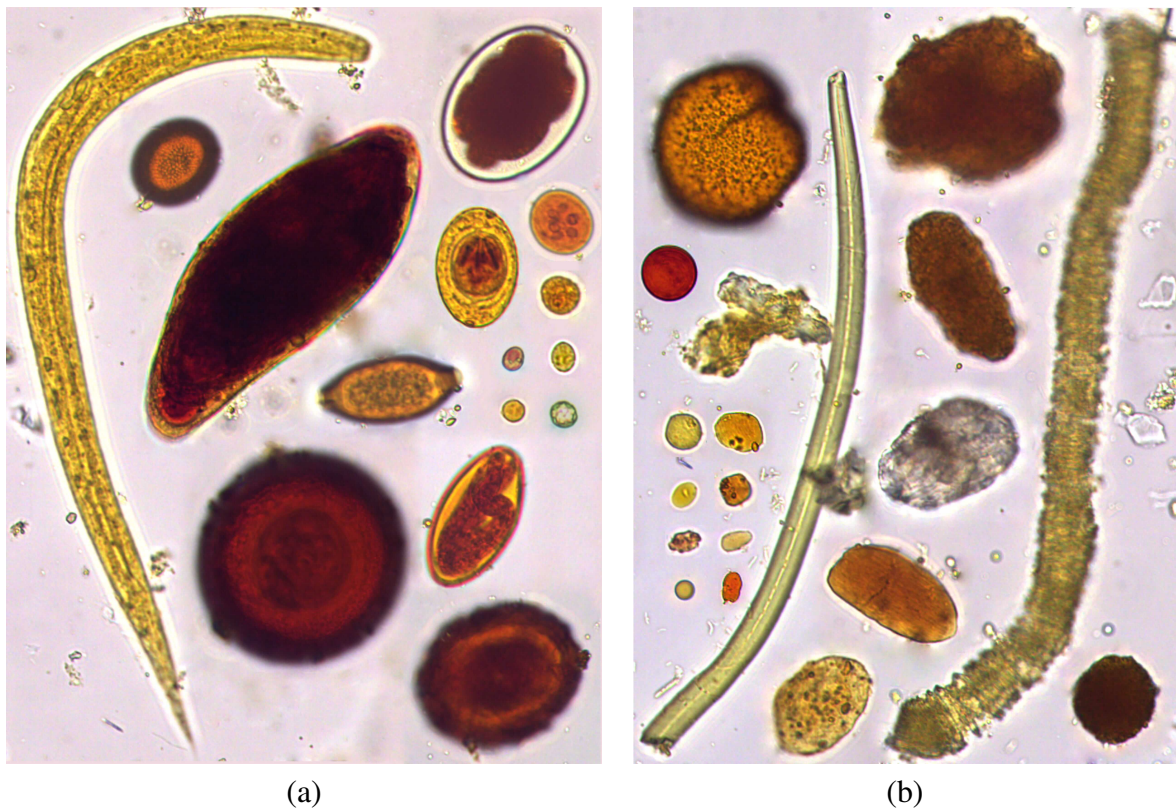


Figura 20: Exemplos de imagens do conjunto de enteroparasitos. (a) espécies de enteroparasitos. (b) impurezas fecais similares às espécies de enteroparasitos.

Tabela 1: Quantidades de classes e imagens em cada um dos subconjuntos de parasitos e impurezas intestinais.

Subconjuntos	Classes	Imagens
EGG-8	8	2.876
EGG-9	9	12.691
EGG-16	16	12.691
LAR-2	2	1.598
PRO-6	6	8.848
PRO-7	7	37.376
PRO-12	12	37.376
PAR-15	15	11.971
PAR-16	16	51.665
PAR-30	30	51.665

Tabela 2: Descrição das classes e distribuição das imagens em cada classe dos subconjuntos de parasitos e impurezas intestinais.

Classes		EGG-8	EGG-9	EGG-16	LAR-2	PRO-6	PRO-7	PRO-12	PAR-15	PAR-16	PAR-30
Ovos	A.lumbricoides	835	835	835	-	-	-	-	835	835	835
	Ancilostomideo	286	286	286	-	-	-	-	286	286	286
	E.vermicularis	103	103	103	-	-	-	-	103	103	103
	H.diminuta	83	83	83	-	-	-	-	83	83	83
	H.nana	501	501	501	-	-	-	-	501	501	501
	S.mansoni	254	254	254	-	-	-	-	254	254	254
	T.trichiura	435	435	435	-	-	-	-	435	435	435
Taenia	379	379	379	-	-	-	-	379	379	379	
Larvas	S.stercoralis	-	-	-	247	-	-	-	247	247	247
Protozoários	B.hominis	-	-	-	-	309	309	309	309	309	309
	E.coli	-	-	-	-	869	869	869	869	869	869
	E.histolytica	-	-	-	-	659	659	659	659	659	659
	E.nana	-	-	-	-	1783	1783	1783	1783	1783	1783
	Giardia	-	-	-	-	1931	1931	1931	1931	1931	1931
	I.butshlii	-	-	-	-	3297	3297	3297	3297	3297	3297
Impurezas Ovos	A.lumbricoides	-	9815	3051	-	-	-	-	-	39694	3051
	Ancilostomideo	-		549	-	-	-	-	-		549
	E.vermicularis	-		446	-	-	-	-	-		446
	H.diminuta	-		233	-	-	-	-	-		233
	H.nana	-		1054	-	-	-	-	-		1054
	S.mansoni	-		1917	-	-	-	-	-		1917
	T.trichiura	-		1360	-	-	-	-	-		1360
Taenia	-	1205	-	-	-	-	-	1205			
Impurezas Larvas	S.stercoralis	-	-	-	1351	-	-	-	-	-	1351
Impurezas Protozoários	B.hominis	-	-	-	-	-	-	3279	-	-	3279
	E.coli	-	-	-	-	-	-	1041	-	-	1041
	E.histolytica	-	-	-	-	-	-	674	-	-	674
	E.nana	-	-	-	-	-	-	6891	-	-	6891
	Giardia	-	-	-	-	-	-	6953	-	-	6953
	I.butshlii	-	-	-	-	-	-	9690	-	-	9690

4 CONTRIBUIÇÃO 1 - DESCRITORES PARA IMAGENS - DEEP FEATURES VS HANDCRAFTED FEATURES

Experimentos foram realizados para análise dos conjuntos de dados e das técnicas adequadas a serem consideradas em cada uma das etapas de classificação e de geração de imagens sintéticas da abordagem proposta no Capítulo 6.

4.1 METODOLOGIA PROPOSTA

Foi desenvolvido um estudo comparativo de desempenhos (valores de acurácias) de diferentes métodos para classificação de folhas de plantas, utilizando características extraídas por meio de extratores tradicionais (*handcrafted features*) e por meio de aprendizado profundo (*deep features*) (ALVES et al., 2017). Tal estudo também gerou contribuições em Bressan et al. (2018) e Valerio et al. (2019). A Figura 21 ilustra o *pipeline* considerado para a análise comparativa.

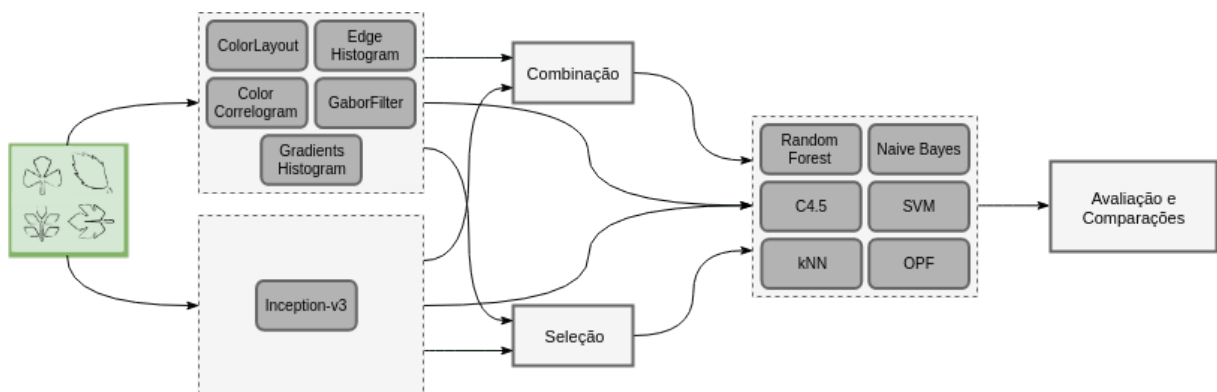


Figura 21: Pipeline da análise comparativa de desempenho, considerando abordagens de aprendizados tradicionais e de aprendizado profundo para classificação de folhas de plantas.

Foram exploradas diferentes técnicas para extração, combinação e seleção de características a partir das imagens. Para extração de características foram adotadas as técnicas sumarizadas na Tabela 3. A primeira coluna corresponde ao nome para cada grupo de extratores. As siglas **C**, **G**, **S** e **T**, referem-se ao tipo de extrator como Cor, Genérico, Forma e Textura, respectivamente. A segunda coluna consiste nos nomes das técnicas utilizadas para extração

das características, e por fim a última coluna denota a quantidade de características extraídas em cada grupo de técnica. Cada conjunto de características são representações numéricas de uma imagem.

Tabela 3: Grupos de técnicas de extração de características considerados nos experimentos e as respectivas quantidades de características obtidas.

Grupos	Técnica	#Features
C_1	<i>ColorLayout MPEG-7</i> (KASUTANI; YAMADA, 2001)	33
C_2	<i>ColorCorrelogram</i> (HUANG et al., 1999)	1.024
G_1	<i>Inception-v3 Transfer Learning</i> (SZEGEDY et al., 2016)	2.048
S_1	<i>Pyramid Histogram of Oriented Gradients</i> (PHOG)(BAI et al., 2009)	630
T_1	<i>Edge Histogram MPEG-7</i> (WON et al., 2002)	80
T_2	<i>Gabor Filter</i> (NIXON; AGUADO, 2012)	60

O grupo de extratores G_1 explora a arquitetura *Inception-v3*, como um extrator genérico de características por meio do uso da estratégia de *Transfer Learning*. Nesse caso, utilizou-se a rede pré-treinada no conjunto de dados da ILSVRC 2012, de forma a obter um vetor com 2.048 características, o qual é a saída da ante-penúltima camada dessa rede. A Figura 22 ilustra a arquitetura *Inception-v3* com ênfase na camada utilizada para a extração das características.

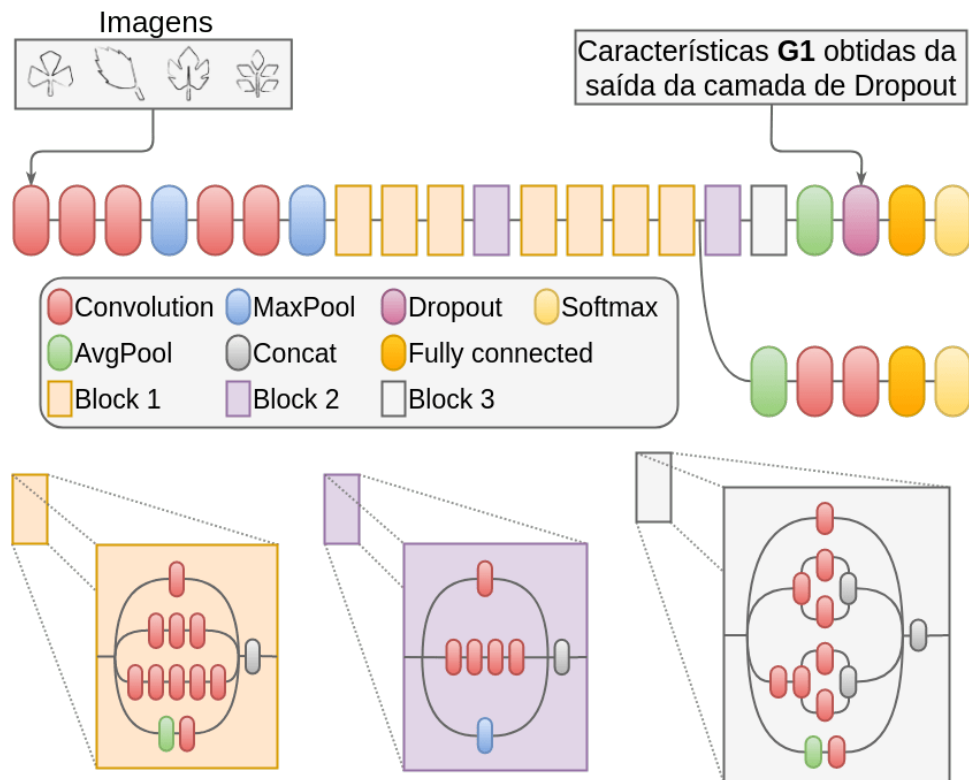


Figura 22: Arquitetura *Inception-v3*, destacando a saída da camada de *dropout* utilizada como vetor de características obtido por G_1 .

A fim de aprofundar a análise e em busca de melhores resultados, foram utilizadas

duas abordagens (CFS+BF e IGR+R) para seleção das melhores características.

- **Correlation-based Feature Selection + Best First (CFS+BF)** (KAREGOWDA et al., 2010): avalia o valor de um subconjunto de características, considerando a habilidade preditiva individual de cada característica juntamente com o grau de redundância entre eles. São preferidos os subconjuntos de características que estão altamente correlacionadas com a classe, apresentando baixa intercorrelação. Inicia com o conjunto vazio e procura o espaço de subconjuntos de características de forma gulosa com *backtracking*. O número de nós consecutivos não melhorados é um parâmetro de parada para o algoritmo.
- **Information Gain Ratio + Ranker (IGR+R)** (KAREGOWDA et al., 2010): avalia o valor de uma característica com base na medida de ganho de informação em relação à cada classe. Classifica as características por meio de suas avaliações individuais seguindo a Equação 7, onde o termo H é a entropia.

$$IGR(Class, Attribute) = \frac{H(Class) - H(Class|Attribute)}{H(Attribute)} \quad (7)$$

A Tabela 4 apresenta cada abordagem de seleção adotada para cada extrator, incluindo a quantidade de características obtidas. Note que algumas abordagens foram omitidas, pois o número de características é um fator limitante. Por exemplo, o extrator C_1 apresenta apenas 33 características, portanto não é possível aplicar as técnicas de Ranking, dado que foram definidos como 50, 100, 200 e 300 as quantidades de características selecionadas.

Tabela 4: Técnicas de seleção e quantidade de características obtidas para cada extrator.

Extrator	CFS+BF	IGR+R-50	IGR+R-100	IGR+R-200	IGR+R-300
C_1	C_1 -CFS+BF-18	-	-	-	-
C_2	C_2 -CFS+BF-103	C_2 -IGR+R-50	C_2 -IGR+R-100	C_2 -IGR+R-200	C_2 -IGR+R-300
G_1	G_1 -CFS+BF-746	G_1 -IGR+R-50	G_1 -IGR+R-100	G_1 -IGR+R-200	G_1 -IGR+R-300
S_1	S_1 -CFS+BF-143	S_1 -IGR+R-50	S_1 -IGR+R-100	S_1 -IGR+R-200	S_1 -IGR+R-300
T_1	T_1 -CFS+BF-37	T_1 -IGR+R-50	-	-	-
T_2	T_2 -CFS+BF-2	T_2 -IGR+R-50	-	-	-

Além da seleção, em alguns casos, a combinação de características pode melhorar os resultados. No entanto, determinar quais as características que devem ser utilizadas em conjunto é um trabalho de experimentação e a melhor combinação geralmente estará associada a um domínio ou conjunto de dados específico. Para tanto, foi realizada a combinação par a par de G_1 com as demais características (C_1 , C_2 , S_1 , T_1 , T_2) seguida da seleção por meio da técnica **CFS+BF**.

Além dos extratores, também foram considerados diferentes algoritmos para a classificação, entre eles: k -NN (COVER; HART, 1967), SVM (HEARST et al., 1998), RF

(BREIMAN, 2001), NB (DOMINGOS; PAZZANI, 1997), C.45 (QUINLAN, 2014) e OPF (PAPA et al., 2009). Cada experimento foi executado 10 vezes e foram utilizados 80% dos dados para treinamento e 20% para teste.

4.2 EXPERIMENTOS E RESULTADOS

As Tabelas 5 e 6 mostram os resultados aplicando-se técnicas de extração de características sem e com seleção das mesmas, respectivamente, para classificação do conjunto de dados de folhas. A Tabela 7 apresenta os resultados com a combinação par a par de G_1 com as demais características (C_1, C_2, S_1, T_1, T_2) seguida da seleção por meio da técnica **CFS+BF**. Os valores sublinhados nas Tabelas representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.

Nos experimentos realizados sem a seleção de características (Tabela 5), o extrator de cor C_2 obteve bons resultados com o classificador OPF, atingindo uma acurácia média de 91.8%. No entanto, observa-se que a extração de características com o método genérico G_1 (*Transfer Learning* na arquitetura *Inception-v3*) se sobressai em todos os algoritmos de classificação avaliados. A melhor acurácia reportada para o extrator G_1 sem seleção foi de 98.48% utilizando o classificador SVM. Analisando o desempenho dos algoritmos de classificação para cada um dos extratores, o classificador OPF, de uma forma geral, apresentou os melhores resultados, atingindo até 96.95% de acurácia utilizando G_1 .

Com a seleção de características (Tabela 6), observa-se algumas melhorias. Por exemplo, nota-se um desempenho discretamente superior com os 746 melhores atributos do grupo G_1 (G_1 -**CFS+BF**-746) quando comparados ao G_1 sem nenhum tipo de seleção (Tabela 5). Além disso, essas discretas melhorias são refletidas para todos os algoritmos de classificação. A melhor acurácia média reportada com a seleção de atributos foi de 98.96% com o uso do classificador SVM e os atributos obtidos por G_1 -**CFS+BF**-746. De forma geral, os melhores resultados foram observados com a técnica de seleção **CFS+BF**.

Nos experimentos envolvendo a combinação de características (Tabela 7), o atributo G_1 foi utilizado como base para combinação par a par com os extratores C_1, C_2, S_1, T_1, T_2 , seguido de seleção por meio da técnica **CFS+BF**. Os resultados mostram discreta melhoria em todos os algoritmos de classificação e combinações quando comparado com os experimentos sem nenhum tipo de combinação. A melhor acurácia média obtida foi com a combinação de G_1 e C_1 ($G_1 + C_1$ -**CFS+BF**-749) com o classificador SVM atingindo 99.30% das amostras

classificadas corretamente.

Tabela 5: Acurácias médias e desvios-padrão obtidos no conjunto PI@ntLeaves II sem seleção de características. Os valores sublinhados representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.

Extrator	NB	SVM	k -NN	C4.5	RF	OPF
C_1	65.5 ± 0.01	66.5 ± 0.01	65.8 ± 0.12	60.3 ± 0.08	75.7 ± 0.02	81.9 ± 0.07
C_2	60.1 ± 0.07	90.5 ± 0.09	86.5 ± 0.10	57.9 ± 0.10	87.6 ± 0.01	91.8 ± 0.03
G_1	<u>90.21 ± 1.20</u>	98.48 ± 0.37	<u>95.31 ± 0.79</u>	<u>62.53 ± 1.96</u>	<u>92.17 ± 0.78</u>	<u>96.95 ± 0.79</u>
S_1	72.6 ± 0.03	91.10 ± 0.06	80.9 ± 0.23	44.5 ± 0.04	83.9 ± 0.03	89.2 ± 0.01
T_1	62.8 ± 0.04	75.9 ± 0.09	67.6 ± 0.09	39.9 ± 0.02	70.6 ± 0.02	80.7 ± 0.10
T_2	21.0 ± 0.08	27.4 ± 0.05	22.6 ± 0.04	25.2 ± 0.02	26.0 ± 0.01	58.3 ± 0.02

Tabela 6: Acurácias médias e desvios-padrão obtidos no conjunto PI@ntLeaves II com seleção das melhores características. Os valores sublinhados representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.

Features	NB	SVM	k -NN	C4.5	RF	OPF
C_1 -CFS+BF-18	64.48 ± 1.29	64.08 ± 1.46	70.20 ± 1.34	60.46 ± 1.55	75.23 ± 1.49	70.67 ± 0.27
C_2 -CFS+BF-103	69.95 ± 1.46	86.34 ± 1.15	86.23 ± 1.18	57.74 ± 1.80	86.94 ± 1.27	85.98 ± 0.10
C_2 -IGR+R-50	13.86 ± 1.09	16.72 ± 0.94	16.97 ± 0.98	17.98 ± 0.98	18.83 ± 0.98	18.28 ± 0.36
C_2 -IGR+R-100	26.82 ± 1.28	36.69 ± 1.30	35.96 ± 1.54	36.75 ± 1.55	42.51 ± 1.63	17.90 ± 0.14
C_2 -IGR+R-200	54.77 ± 1.49	77.84 ± 1.20	77.55 ± 1.41	54.95 ± 1.84	80.89 ± 1.40	35.25 ± 0.94
C_2 -IGR+R-300	59.71 ± 1.87	86.56 ± 1.08	83.75 ± 1.37	57.40 ± 1.74	85.87 ± 1.17	73.58 ± 0.49
G_1 -CFS+BF-746	<u>92.58 ± 0.87</u>	98.96 ± 0.38	<u>96.89 ± 0.65</u>	<u>63.11 ± 2.03</u>	<u>93.86 ± 0.62</u>	<u>95.89 ± 0.90</u>
G_1 -IGR+R-50	77.60 ± 1.41	88.29 ± 1.01	<u>84.24 ± 1.39</u>	55.80 ± 1.81	85.43 ± 1.12	88.40 ± 0.13
G_1 -IGR+R-100	83.46 ± 1.32	95.04 ± 0.76	91.01 ± 1.05	59.43 ± 1.56	89.37 ± 0.98	82.21 ± 0.53
G_1 -IGR+R-200	85.96 ± 1.31	97.68 ± 0.53	93.93 ± 0.83	60.80 ± 1.46	91.27 ± 0.91	92.15 ± 0.87
G_1 -IGR+R-300	87.31 ± 1.17	98.29 ± 0.42	95.28 ± 0.75	61.57 ± 1.80	92.02 ± 0.81	88.31 ± 0.21
S_1 -CFS+BF-143	72.77 ± 1.47	83.69 ± 1.32	77.16 ± 1.48	45.74 ± 1.85	79.86 ± 1.44	83.77 ± 0.73
S_1 -IGR+R-50	43.33 ± 1.19	59.53 ± 1.61	50.82 ± 1.33	38.25 ± 1.75	59.76 ± 1.61	74.16 ± 0.11
S_1 -IGR+R-100	54.68 ± 1.31	73.85 ± 1.45	62.40 ± 1.41	42.56 ± 1.79	69.61 ± 1.52	56.32 ± 0.30
S_1 -IGR+R-200	63.64 ± 1.27	81.85 ± 1.28	72.62 ± 1.52	46.17 ± 1.77	76.65 ± 1.41	67.96 ± 0.50
S_1 -IGR+R-300	67.89 ± 1.41	84.39 ± 1.19	75.38 ± 1.52	45.97 ± 1.71	78.77 ± 1.56	72.96 ± 0.98
T_1 -CFS+BF-37	65.49 ± 1.59	71.34 ± 1.38	60.79 ± 1.53	40.32 ± 1.32	67.93 ± 1.40	70.73 ± 0.94
T_1 -IGR+R-50	61.46 ± 1.39	70.63 ± 1.58	60.39 ± 1.49	40.19 ± 1.57	66.89 ± 1.63	63.29 ± 0.05
T_2 -CFS+BF-2	25.75 ± 1.13	17.72 ± 0.67	23.53 ± 1.29	24.96 ± 1.37	24.92 ± 1.18	25.44 ± 0.07
T_2 -IGR+R-50	21.32 ± 0.94	28.98 ± 1.15	23.51 ± 1.26	24.99 ± 1.35	25.21 ± 1.40	26.74 ± 0.73

Tabela 7: Acurácias médias e desvios-padrão obtidos no conjunto PI@ntLeaves II combinando as melhores características de G_1 . Os valores sublinhados representam os melhores resultados para cada algoritmo de classificação. Os valores em negrito correspondem aos melhores resultados obtidos para cada extrator de características.

Features	NB	SVM	k -NN	C4.5	RF	OPF
G_1+C_1 -CFS+BF-749	<u>93.33 ± 0.44</u>	99.30 ± 0.65	<u>97.27 ± 0.75</u>	63.48 ± 0.33	94.09 ± 0.81	96.07 ± 0.02
G_1+C_2 -CFS+BF-814	93.25 ± 0.94	99.06 ± 0.52	<u>97.12 ± 0.56</u>	63.45 ± 0.05	93.92 ± 0.44	96.29 ± 0.31
G_1+S_1 -CFS+BF-865	92.63 ± 0.19	99.08 ± 0.44	97.03 ± 0.28	63.43 ± 0.65	<u>94.26 ± 0.89</u>	96.12 ± 0.51
G_1+T_1 -CFS+BF-774	92.68 ± 0.97	99.01 ± 0.94	97.07 ± 0.30	<u>63.56 ± 0.65</u>	<u>93.86 ± 0.27</u>	<u>96.36 ± 0.62</u>
G_1+T_2 -CFS+BF-746	92.58 ± 0.29	99.02 ± 0.33	96.89 ± 0.08	<u>63.11 ± 0.93</u>	93.86 ± 0.92	95.89 ± 0.01

Experimentos também foram realizados com os conjuntos de dados de enteroparasitos. Nesse caso, foram utilizadas duas redes neurais profundas, a *Inception-v3* e *ResNet18*,

com o intuito de adquirir um *baseline* para comparações com as demais estratégias de aprendizado (Capítulo 5). Para treinamento das redes foram definidos os parâmetros: estratégia de otimização *Adam* (KINGMA; BA, 2014); taxa de aprendizado 0.01; tamanho do *batch* 20; e número de épocas 50. Os resultados são apresentados na Tabela 8. As arquiteturas *ResNet18* e *Inception-v3* apresentaram resultados similares para todos os subconjuntos de dados de enteroparasitos e impurezas intestinais.

Tabela 8: Acurácias médias e desvios-padrão obtidos pelas arquiteturas *ResNet18* e *Inception-v3* nos conjuntos de dados de enteroparasitos.

Conjunto	ResNet18	Inception-v3
EGG-8	99.00 ± 1.01	98.18 ± 1.01
EGG-9	96.46 ± 1.34	97.01 ± 0.62
EGG-16	71.87 ± 2.41	72.56 ± 3.46
LAR-2	93.03 ± 1.68	95.05 ± 1.60
PRO-6	84.65 ± 2.45	77.91 ± 3.70
PRO-7	89.07 ± 1.97	87.38 ± 1.93
PRO-12	60.43 ± 3.53	57.34 ± 4.08
PAR-15	87.76 ± 1.64	83.71 ± 3.26
PAR-16	90.15 ± 2.41	89.55 ± 1.32
PAR-30	61.65 ± 3.88	60.75 ± 4.25

Técnicas de redução de dimensionalidade como *Principal Components Analysis* (PCA) são amplamente utilizadas e bem sucedidas (BRYANT; YARNOLD, 1995; PRICE et al., 2006; DU; ZHU, 2019). Sendo assim, a técnica de redução PCA foi avaliada, com o intuito de reduzir a complexidade (sem afetar o desempenho na classificação) das técnicas de aprendizado consideradas nos demais experimentos (Capítulos 5 e 6). As características obtidas a partir da *Inception-v3* foram submetidas a uma redução incremental da quantidade de dimensões (iniciando com tamanho 2.048), de forma a encontrar um ponto ótimo mínimo para a dimensionalidade. A Figura 23 ilustra o impacto nos valores de acurácia a medida que a dimensão (tamanho) do vetor de características diminui.

Para cada tamanho de dimensionalidade analisado, foram realizadas 5 execuções, com separação de 80% dos dados para treinamento e 20% para testes. A classificação foi realizada com o algoritmo SVM para os conjunto de dados LEA-53, PAR-15, PAR-16 e CIFAR-10. Considerando os conjuntos LEA-53, PAR-15 e PAR-16 pode-se verificar o ponto ótimo mínimo por volta da dimensão 120, ou seja, em que obtém-se o maior valor de acurácia com a menor quantidade de características. Não houve variações significantes de acurácia para o conjunto CIFAR-10. Acredita-se que esse fato deve-se à baixa qualidade (resolução) das imagens desse conjunto.

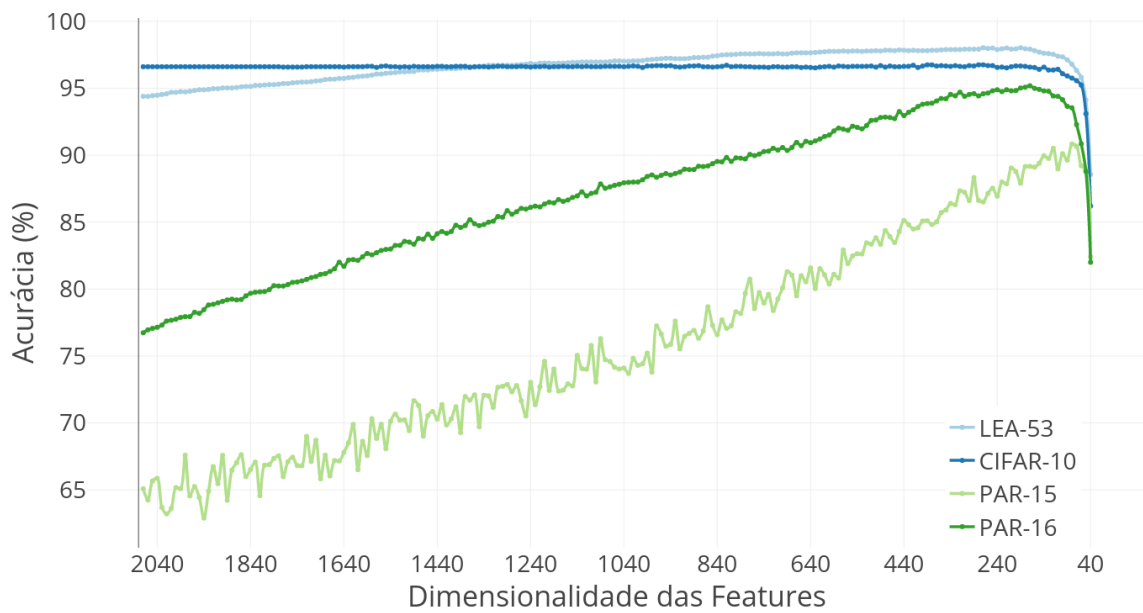


Figura 23: Acurácias médias obtidas com RF a partir das reduções de dimensionalidade das características (Inception-v3) utilizando PCA para cada conjunto de dados (LEA-53, CIFAR-10, PAR-15 e PAR-16).

5 CONTRIBUIÇÃO 2 - NOVAS ESTRATÉGIAS DE APRENDIZADO ATIVO

Para melhorar os processos de sintetização das imagens e de aprendizado do classificador é proposta uma nova variação na inicialização das estratégias clássicas de aprendizado ativo (*Entropy*, *Least Confidence* e *Margin Sampling*), descrita na Seção 5.1. Também são propostas estratégias de aprendizado ativo (ALVES et al., 2019b), as quais são descritas nas Seções 5.3-5.4.

5.1 ENTROPY CLUSTER (EN-CLU), LEAST CONFIDENCE CLUSTER (LC-CLU) E MARGIN SAMPLING CLUSTER (MS-CLU)

Considerando as técnicas tradicionais de aprendizado ativo *Entropy* (EN), *Least Confidence* (LC) e *Margin Sampling* (MS) (Subseção 2.2.1), foi proposta uma nova variação para seleção das amostras que compõem a primeira iteração do ciclo de aprendizado. A seleção inicial, consiste na utilização de uma técnica de *clustering* no conjunto de dados (inicialmente não rotulado), obtendo as amostras raízes de cada *cluster*.

Em contraste com a seleção inicial aleatória, propostas nos métodos clássicos originais, essa nova variação explora a diversidade dos dados, fazendo com que, logo na primeira iteração, o classificador aprenda com um conjunto de amostras mais informativas (representativas), pois há maior probabilidade de que amostras de classes distintas estarão presentes nesse grupo inicial de amostras.

Tais técnicas foram denominadas como *Entropy Cluster* (EN-CLU), *Least Confidence Cluster* (LC-CLU) e *Margin Sampling Cluster* (MS-CLU). A Figura 24 ilustra a composição das amostras na primeira iteração para essas estratégias.

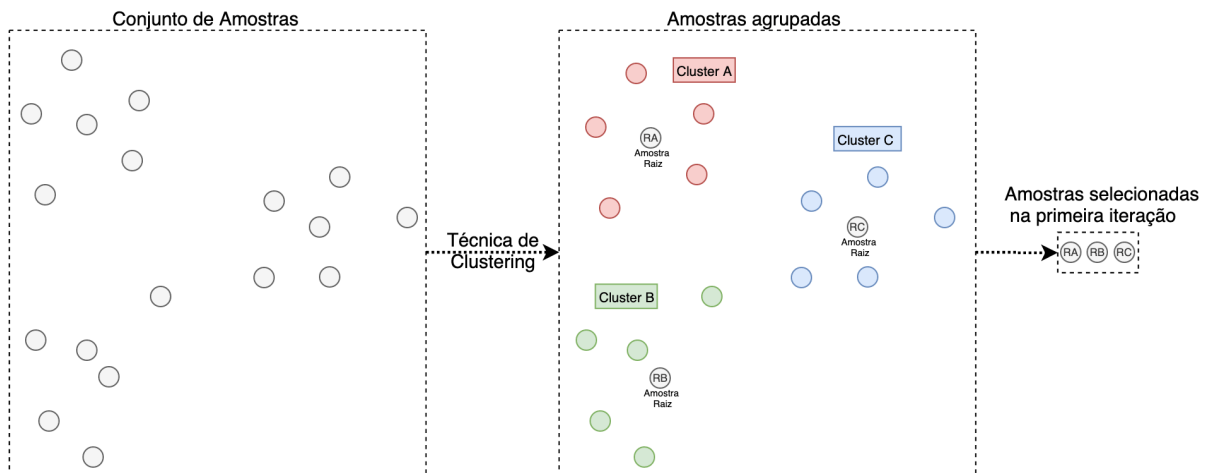


Figura 24: Amostras obtidas na primeira iteração pelas estratégias EN-CLU, LC-CLU e MS-CLU.

5.2 MINIMUM-SPANNING TREE CLUSTER DECREASING DISAGREE EDGES (MST-CLU-DDE)

A técnica *Minimum-Spanning Tree Cluster Decreasing Disagree Edges* (MST-CLU-DDE) está inserida no grupo de estratégias de aprendizado ativo que realiza pré-processamento das amostras, organizando-nas para posterior apresentação ao classificador a cada iteração do aprendizado. Portanto, essa técnica também é adequada para aplicações que requerem tempo de resposta interativo, pois ao contrário das técnicas clássicas (*Entropy*, *Least Confidence* e *Margin Sampling*), evita que todo conjunto de dados seja reprocessado a cada iteração do aprendizado.

A estratégia MST-CLU-DDE define duas macro etapas de organização e de seleção dos dados. A etapa de organização da MST-CLU-DDE é iniciada a partir do agrupamento do conjunto de dados não rotulado. A Figura 25 ilustra um exemplo da etapa de organização da técnica MST-CLU-DDE considerando 2 clusters (A e B). O número de clusters e a técnica utilizada para realizar o agrupamento não são fixos e podem ser parametrizados. No entanto, para otimizar a diversidade das amostras, é recomendado que o número de clusters seja maior ou igual ao número de classes distintas presentes no conjunto avaliado. Para cada cluster, são obtidas e separadas as amostras raízes. Em seguida, também para cada cluster, sem a presença das amostras raízes, aplica-se o algoritmo da *Minimum Spanning Tree* (MST) seguido da ordenação decrescente (de pesos das arestas) das amostras. Então, são obtidas listas ordenadas para cada cluster. Na sequência, uma lista final é construída com os elementos obtidos de forma: sequencial, intercalada e em pares a partir de cada lista ordenada anteriormente.

A etapa de seleção da MST-CLU-DDE inicia com o conjunto de amostras raízes e a lista final, obtidos como saída da etapa de organização. A primeira instância do classificador é obtida com o treinamento a partir do conjunto de amostras raízes. Nas iterações seguintes, a

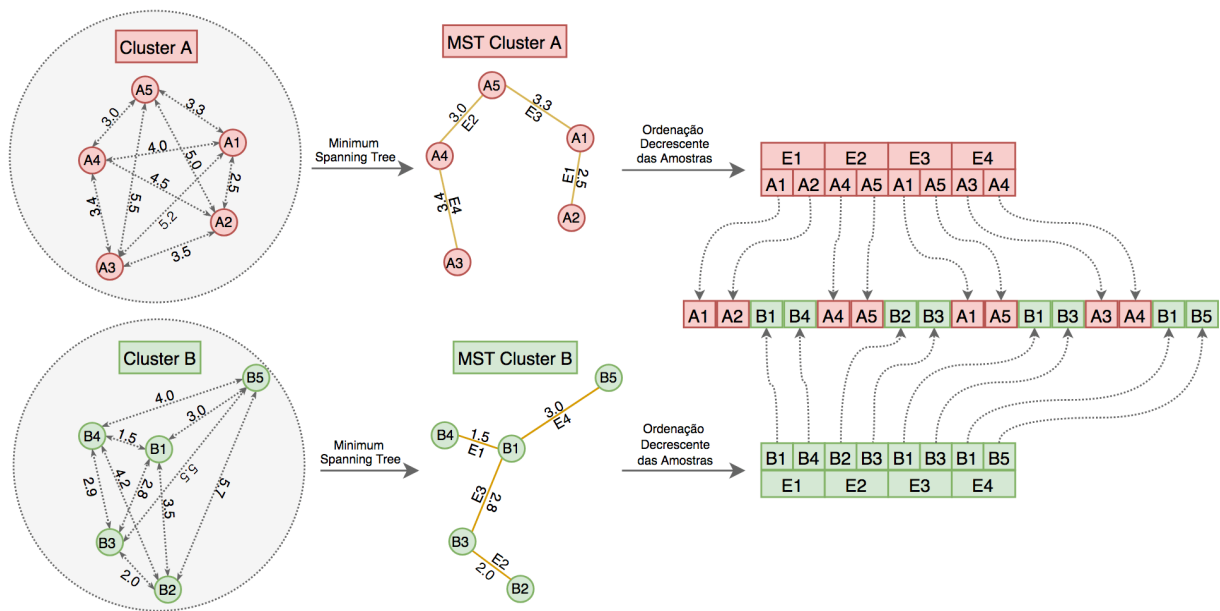


Figura 25: Etapa de organização da estratégia de aprendizado ativo *Minimum Spanning Tree Cluster Decreasing Disagree Edges* (MST-CLU-DDE).

instância atual do classificador será consultada para rotular uma aresta (par de amostras) por vez e auxiliar no processo de seleção. Se as amostras apresentam rótulos distintos, bem como ainda não foram selecionadas e o limite de amostras a cada iteração ainda não foi atingido, estas serão selecionadas e exibidas ao especialista para anotação. Esse comportamento se repete, avaliando cada aresta da lista final ordenada, até que o limite de amostras selecionadas a cada iteração seja alcançado. Caso nenhum par (aresta) de amostras restante na lista satisfaça os critérios para seleção, o primeiro par de amostras remanescente é selecionado.

A cada iteração do aprendizado, o classificador aprende e participa do seu processo de aprendizado, classificando e auxiliando na seleção de amostras mais informativas. Diferentemente de abordagens clássicas, MST-CLU-DDE não requer a classificação de todas as amostras a cada iteração do aprendizado. Nesse caso, para seleção das amostras mais informativas, um pequeno conjunto de amostras é obtido e classificado.

MST-CLU-DDE explora diferentes critérios de seleção. Critérios com base em **representatividade** ao realizar o agrupamento das amostras e obter as raízes (amostras representativas) dos grupos. Critérios de **diversidade** ao obter listas diversas compostas por amostras correspondentes a cada grupo. Critérios de **incerteza** ao priorizar amostras oriundas das arestas (máximas das mínimas) das MSTs, ou seja, amostras que, embora sejam as mais próximas, possivelmente sejam de classes distintas. Além disso, no MST-CLU-DDE, o critério de incerteza é fortemente considerado ao explorar o classificador a cada iteração do processo de aprendizado, auxiliando e selecionando as amostras mais incertas (rotuladas em classes distintas).

O Algoritmo 1 define a etapa de organização da estratégia MST-CLU-DDE que inicia com a divisão do conjunto de dados \mathcal{L} em \mathcal{L}_2 e \mathcal{L}_3 referentes aos conjuntos de aprendizado e de teste. Como próximo passo, realiza-se o agrupamento das amostras do conjunto de aprendizado \mathcal{L}_2 (Linha 2) e obtém-se as raízes \mathcal{R} dos *clusters* $\mathcal{C}_i, i = 1, 2, 3 \dots k$ (Linha 3). Então, para cada *cluster* \mathcal{C}_i (Linha 4), é calculada uma *Minimum Spanning Tree* (MST) considerando amostras do respectivo *cluster* (Linha 5). Em seguida, as arestas da MST são ordenadas de forma decrescente (Linha 5) para cada *cluster* \mathcal{C}_i , gerando as listas ordenadas \mathcal{L}_i . A etapa de organização é finalizada com o retorno de uma lista organizada \mathcal{F} (Linha 13) construída a partir das amostras (arestas) intercaladas de cada lista \mathcal{L}_i (Linhas 8-12).

Algoritmo 1: Estratégia de Organização - MST-CLU-DDE

Entrada: conjunto de aprendizado não anotado \mathcal{L}_2 .

Saída: conjunto de aprendizado organizado \mathcal{F} e conjunto de raízes \mathcal{R} .

Dados: conjuntos \mathcal{C}_i com amostras de cada *cluster* $i = 1, 2, \dots, k$ e o número de *clusters* k .

```

1 início
2    $\mathcal{C}_i, i = 1, 2, \dots, k \leftarrow$  computa clusters de  $\mathcal{L}_2$ ;
3    $\mathcal{R} \leftarrow$  raízes dos clusters de  $\mathcal{L}_2$ ;
4   para cada  $\mathcal{C}_i, i = 1, 2, \dots, k$  faça
5     |  $\mathcal{L}_i \leftarrow$  MST( $\mathcal{C}_i$ );
6     |  $\mathcal{L}_i \leftarrow$  Sort( $\mathcal{L}_i$ );
7   fim
8   repita
9     | para cada  $\mathcal{L}_i, i = 1, 2, \dots, k$  faça
10    | |  $\mathcal{F} \leftarrow \mathcal{F} \cup$  NextPair( $\mathcal{L}_i$ );
11    | fim
12  até  $\mathcal{L}_i$  ainda possua amostras;
13  retorna  $\mathcal{F}$  e  $\mathcal{R}$ 
14 fim

```

Para casos em que os conjuntos considerados são grandes ou apresentam dimensões elevadas para descrição das amostras, é possível otimizar o tempo de treinamento paralelizando, na etapa de organização, as operações de cálculo e ordenação da MST (Linhas 4-7) para cada *cluster*.

O Algoritmo 2 define a etapa de seleção da estratégia MST-CLU-DDE. Na primeira iteração são selecionadas as amostras raízes dos *clusters* (Linha 2). Na sequência, o especialista anota as amostras raízes (Linha 3), gerando a primeira instância do classificador. Nas iterações seguintes, de forma sequencial e em pares, amostras da lista previamente organizada são avaliadas. Se os pares de amostras são rotulados em classes distintas, de acordo com o rótulo fornecido pela instância atual do classificador, estes serão selecionados. Se tal critério

não for satisfeito, o próximo par de amostras da lista é analisado até que não existam mais pares de amostras disponíveis para serem avaliados. Por fim, o especialista corrige os rótulos das amostras e o processo se repete até que o mesmo esteja satisfeito (Linhas 4-9).

Algoritmo 2: Estratégia de Seleção - MST-CLU-DDE

Entrada: conjunto de aprendizado organizado \mathcal{F} e conjunto de raízes \mathcal{R} .
Saída: classificador treinado.
Dados: conjunto de treinamento \mathcal{L}_1 e conjunto de amostras selecionado \mathcal{L}'_1 .

- 1 **início**
- 2 $\mathcal{L}_1 \leftarrow \mathcal{R}$;
- 3 Especialista anota as amostras em \mathcal{L}_1 ;
- 4 **repita**
- 5 (Re)-treinar o classificador com \mathcal{L}_1 ;
- 6 $\mathcal{L}'_1 \leftarrow$ novas amostras classificadas em classes distintas seguindo a ordem da lista pré-organizada \mathcal{F} ;
- 7 Especialista corrige os rótulos das amostras em \mathcal{L}'_1 ;
- 8 $\mathcal{L}_1 \leftarrow \mathcal{L}_1 \cup \mathcal{L}'_1$;
- 9 **até que especialista satisfeito;**
- 10 **fim**

5.3 MINIMUM-SPANNING TREE CLUSTER DECREASING SAMPLING (MST-CLU-DS)

Como em qualquer estratégia de aprendizado ativo, na *Cluster Minimum-Spanning Tree Decreasing Sampling* (MST-CLU-DS) o intuito é que as amostras mais informativas sejam selecionadas durante os ciclos iniciais de aprendizagem.

Essa técnica também faz parte do grupo de estratégias de aprendizado ativo que realiza a organização das amostras como etapa de pré-processamento, evitando que todo o conjunto de dados seja avaliado a cada iteração do aprendizado. Como consequência tem-se a redução do custo computacional quando comparado a técnicas clássicas da literatura.

A etapa de organização da MST-CLU-DS é semelhante a da MST-CLU-DDE. Assim como na estratégia MST-CLU-DDE, a utilização do conjunto de raízes e do algoritmo MST em cada *cluster*, possibilitam explorar, de forma eficaz, critérios baseados em representatividade, incerteza e diversidade das imagens. Uma das diferenças entre as duas técnicas nessa etapa é que na MST-CLU-DS a lista final é obtida intercalando apenas uma amostra de cada lista (i.e. de cada *cluster*). Na MST-CLU-DDE intercala-se as amostras em pares (arestas das MSTs). A Figura 26 ilustra um exemplo da etapa de organização da técnica MST-CLU-DS considerando 2 *clusters* (A e B).

A principal diferença entre as estratégias MST-CLU-DS e MST-CLU-DDE está na

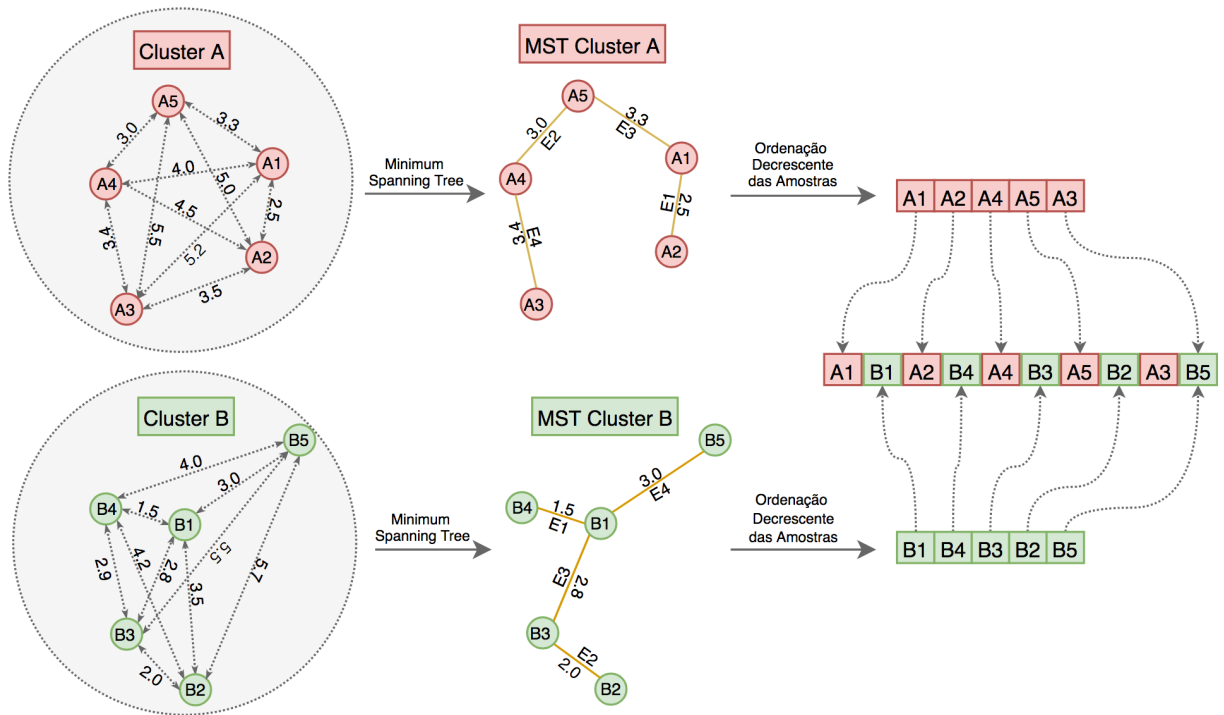


Figura 26: Etapa de organização da estratégia de aprendizado ativo *Minimum-Spanning Tree Cluster Decreasing Sampling* (MST-CLU-DS).

etapa de seleção. No MST-CLU-DDE, a cada iteração, algumas amostras precisam ser avaliadas pela instância atual do classificador antes de serem selecionadas/apresentadas ao especialista, ou seja, o classificador auxilia na tomada de decisão. Isso não ocorre no caso da MST-CLU-DS que obtém as amostras de forma sequencial a partir da lista pré-organizada.

Por essa razão, em termos de tempo de seleção, a estratégia MST-CLU-DS torna-se consideravelmente mais eficiente em relação à estratégia MST-CLU-DDE, podendo atingir um fator de melhoria de 100 vezes mais rápida. Em contrapartida, a MST-CLU-DDE pode obter resultados de acurácia melhores com número reduzido de iterações.

Após a seleção, o classificador pode ser aplicado de forma a previamente rotular o conjunto a ser exibido ao especialista. Sendo assim, diferentemente de abordagens clássicas, MST-CLU-DS não requer a classificação de todas as amostras a cada iteração do aprendizado.

O Algoritmo 3 define os passos da etapa de organização da estratégia MST-CLU-DS. Inicialmente, tem-se a divisão do conjunto de dados original \mathcal{L} em \mathcal{L}_2 e \mathcal{L}_3 , que compõem os conjuntos de aprendizado e de teste, respectivamente. Realiza-se o agrupamento do conjunto de dados de aprendizado \mathcal{L}_2 (Linha 2) (por exemplo, com a técnica *k-means*), obtendo-se as amostras raízes \mathcal{R} (Linha 3) a partir de cada *cluster* $\mathcal{C}_i, i = 1, 2, 3 \dots k$. Em seguida, é calculada a *Minimum-Spanning Tree* (MST) para cada *cluster* \mathcal{C}_i (Linha 4). A partir dos resultados das MSTs aplicadas a cada *cluster*, são criadas e organizadas listas $\mathcal{L}_i, i = 1, 2, 3 \dots, k$ (Linha

5). Cada lista \mathcal{L}_i é composta das amostras do seu respectivo *cluster* e ordenada de forma decrescente de peso das arestas resultantes da MST (Linha 6). Por fim, uma lista final \mathcal{F} é construída, concatenando-se, alternadamente, elemento a elemento das listas $\mathcal{L}_1, \mathcal{L}_2, \dots, \mathcal{L}_k$ (Linhas 8-12). Como saídas da etapa de organização têm-se a lista \mathcal{F} de amostras de aprendizado pré-organizada e o conjunto de amostras raízes \mathcal{R} (linha 13).

Algoritmo 3: Estratégia de Organização - MST-CLU-DS

Entrada: conjunto de aprendizado não anotado \mathcal{Z}_2 .

Saída: conjunto de aprendizado organizado \mathcal{F} e o conjunto de raízes \mathcal{R} .

Dados: conjuntos \mathcal{C}_i com amostras de cada *cluster* $i = 1, 2, \dots, k$ e o número de *clusters* k .

```

1 início
2    $\mathcal{C}_i, i = 1, 2, \dots, k \leftarrow$  computa clusters de  $\mathcal{Z}_2$ ;
3    $\mathcal{R} \leftarrow$  raízes dos clusters de  $\mathcal{Z}_2$ ;
4   para cada  $\mathcal{C}_i, i = 1, 2, \dots, k$  faça
5     |    $\mathcal{L}_i \leftarrow$  MST( $\mathcal{C}_i$ );
6     |    $\mathcal{L}_i \leftarrow$  Sort( $\mathcal{L}_i$ );
7   fim
8   repita
9     |   para cada  $\mathcal{L}_i, i = 1, 2, \dots, k$  faça
10    |   |  $\mathcal{F} \leftarrow \mathcal{F} \cup$  NextSample( $\mathcal{L}_i$ );
11    |   fim
12  até  $\mathcal{L}_i$  ainda possua amostras;
13  retorna  $\mathcal{F}$  e  $\mathcal{R}$ 
14 fim

```

A etapa de seleção é definida no Algoritmo 4. A partir dos dados de saída da etapa de organização, obtém-se o conjunto das amostras raízes \mathcal{R} e a lista ordenada \mathcal{F} . É definido como \mathcal{Z}_1 o conjunto de amostras para treinamento do classificador a cada iteração do aprendizado. O método inicia com \mathcal{Z}_1 recebendo as amostras do conjunto de raízes \mathcal{R} (Linha 2) para compor o conjunto de treinamento da primeira iteração. O especialista anota tais amostras (Linha 3) e é realizado o processo de aprendizado (Linhas 4-9), incluindo o (re-)treinamento das instâncias do classificador (Linha 5). A cada iteração do processo de aprendizado, novas amostras são obtidas a partir da lista pré-organizada \mathcal{F} (Linha 6). O especialista corrige os rótulos das amostras classificadas incorretamente (Linha 7). Sendo assim, as amostras selecionadas são incluídas em \mathcal{Z}_1 (Linha 8). O ciclo se repete até que o especialista esteja satisfeito com o desempenho do classificador (Linhas 4-9).

Algoritmo 4: Estratégia de Seleção - MST-CLU-DS

Entrada: conjunto de aprendizado organizado \mathcal{F} e conjunto de raízes \mathcal{R} .

Saída: classificador treinado.

Dados: conjunto de treinamento \mathcal{L}_1 e conjunto de amostras selecionado \mathcal{L}'_1 .

```

1 início
2    $\mathcal{L}_1 \leftarrow \mathcal{R}$ ;
3   Especialista anota as amostras em  $\mathcal{L}_1$ ;
4   repita
5     (Re)-treinar o classificador com  $\mathcal{L}_1$ ;
6      $\mathcal{L}'_1 \leftarrow$  novas amostras sequencialmente da lista pré-organizada  $\mathcal{F}$ ;
7     Especialista corrige os rótulos das amostras em  $\mathcal{L}'_1$ ;
8      $\mathcal{L}_1 \leftarrow \mathcal{L}_1 \cup \mathcal{L}'_1$ ;
9   até que especialista satisfeito;
10 fim
  
```

5.4 MINIMUM-SPANNING TREE CLUSTER ROOT DISAGREE SAMPLING (MST-CLU-RDS)

A técnica *Minimum-Spanning Tree Cluster Root Disagree Sampling* (MST-CLU-RDS), também realiza pré-processamento das amostras, organizando-nas para posterior apresentação ao classificador a cada iteração do aprendizado. Portanto, essa técnica também é adequada para aplicações que requerem tempo de resposta interativo, dado que a técnica não requer a re-organização e a classificação de todas as amostras do conjunto a cada iteração do aprendizado.

Essa técnica apresenta as etapas de organização e seleção como uma combinação de ideias das estratégias MST-CLU-DDE e RDS (Subseção 2.2.6). A estratégia MST-CLU-RDS, assim como em MST-CLU-DDE, aplica em cada *cluster*, durante a etapa de organização, o algoritmo da MST seguido da ordenação decrescente das amostras. Considerando a etapa de seleção, assim como na estratégia RDS, em MST-CLU-RDS são utilizadas as amostras raízes dos *clusters* para compor os elementos da primeira iteração. Posteriormente, com a informação dos rótulos das amostras raízes, são verificadas quais amostras das listas previamente organizadas possuem classes distintas da amostra raiz a partir do seu respectivo *cluster*.

A Figura 27 ilustra um exemplo da etapa de organização da técnica MST-CLU-RDS considerando 2 *clusters* (A e B). A partir do agrupamento e da aplicação do algoritmo da MST para cada cluster são obtidas as raízes e listas de amostras (arestas) em ordem decrescente. Uma lista final é construída intercalando, a partir de cada *cluster*, o par contendo uma amostra raiz e uma amostra da lista previamente ordenada de um dado *cluster*. Na etapa de seleção, a cada iteração, o classificador será consultado para rotular os pares de amostras da lista final. Caso as amostras de um dado par recebam rótulos distintos, estas serão selecionadas. Caso contrário, o

primeiro par de amostras remanescente é selecionado. Esse processo é repetido até que o limite de amostras selecionadas a cada iteração seja alcançado.

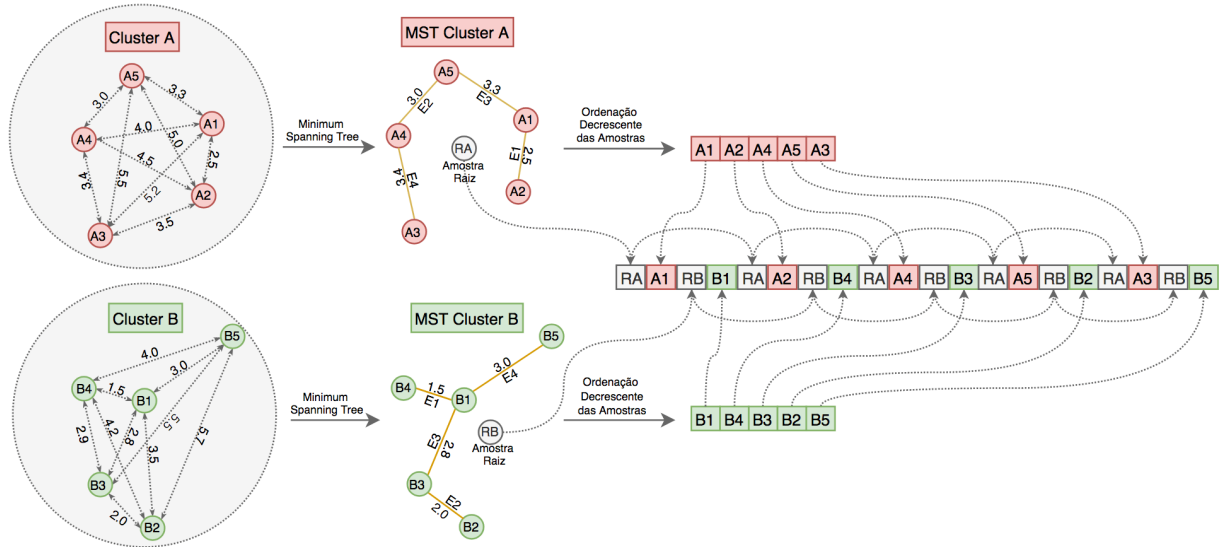


Figura 27: Etapa de organização da estratégia de aprendizado ativo *Minimum-Spanning Tree Cluster Root Disagree Sampling* (MST-CLU-RDS).

Considerando detalhes de implementação, a proposta MST-CLU-RDS apresenta custo computacional menor em relação à estratégia RDS. A MST-CLU-RDS fornece ao final da etapa de organização uma única lista, enquanto que a RDS obtém k listas, em que k representa o número de *clusters*. Em termos de custo computacional na etapa de seleção das amostras, a estratégia MST-CLU-RDS é equivalente à estratégia MST-CLU-DDE. Nessa etapa, em ambas estratégias, apenas algumas amostras precisam ser avaliadas pela instância atual do classificador antes de serem selecionadas/apresentadas ao especialista.

O Algoritmo 5 define a etapa de organização da estratégia MST-CLU-RDS que inicia com a divisão do conjunto de dados \mathcal{L} em \mathcal{L}_2 e \mathcal{L}_3 referente aos conjuntos de treinamento e teste. Como próximo passo, realiza-se o agrupamento das amostras do conjunto de aprendizado \mathcal{L}_2 (Linha 2) e obtém as raízes \mathcal{R} dos *clusters* $\mathcal{C}_i, i = 1, 2, 3 \dots k$ (Linha 3). Em seguida, é calculada a *Minimum Spanning Tree* (MST) para cada *cluster* \mathcal{C}_i (Linha 4) e as amostras das MSTs (Linha 5) são ordenadas de forma decrescente de pesos das arestas (Linha 6). O conjunto de aprendizado é organizado, gerando-se uma lista \mathcal{F} a partir das amostras raízes \mathcal{R}_i e amostras de cada lista ordenada \mathcal{L}_i (Linhas 8-12). A etapa de organização finaliza com o retorno de uma lista de amostras de aprendizado pré-organizada \mathcal{F} e o conjunto de raízes \mathcal{R} (Linha 13).

O Algoritmo 6 define a etapa de seleção da estratégia MST-CLU-RDS. Na primeira iteração são selecionadas as amostras raízes dos *clusters* \mathcal{R} (Linha 2), o especialista anota essas amostras (Linha 3) e treina-se a primeira instância do classificador (Linha 5). São selecionadas

Algoritmo 5: Estratégia de Organização - MST-CLU-RDS

Entrada: conjunto de aprendizado não anotado \mathcal{L}_2 .

Saída: conjunto de aprendizado organizado \mathcal{F} e o conjunto de raízes \mathcal{R} .

Dados: conjuntos \mathcal{C}_i com amostras de cada *cluster* $i = 1, 2, \dots, k$ e o número de *cluster* k .

```

1 início
2    $\mathcal{C}_i, i = 1, 2, \dots, k \leftarrow$  computa clusters de  $\mathcal{L}_2$ ;
3    $\mathcal{R} \leftarrow$  raízes dos clusters de  $\mathcal{L}_2$ ;
4   para cada  $\mathcal{C}_i, i = 1, 2, \dots, k$  faça
5     |    $\mathcal{L}_i \leftarrow$  MST( $\mathcal{C}_i$ );
6     |    $\mathcal{L}_i \leftarrow$  Sort( $\mathcal{L}_i$ );
7   fim
8   repita
9     |   para cada  $\mathcal{L}_i, i = 1, 2, \dots, k$  faça
10    |   |    $\mathcal{F} \leftarrow \mathcal{F} \cup \mathcal{R}_i \cup$  NextSample( $\mathcal{L}_i$ );
11    |   fim
12  até  $\mathcal{L}_i$  ainda possua amostras;
13  retorna  $\mathcal{F}$  e  $\mathcal{R}$ 
14 fim

```

amostras a partir da lista pré-organizada \mathcal{F} (Linha 6), que considera uma amostra de cada *cluster*, seguindo a ordem conforme organização das arestas das MSTs. Caso seja verificado que a amostra apresenta rótulo (fornecido pela instância atual do classificador) diferente do rótulo da amostra raiz do respectivo *cluster*, a amostra deve ser selecionada. Caso contrário a próxima amostra da lista é avaliada. Se as amostras da lista são avaliadas e tal critério de seleção não é satisfeito, a primeira amostra da lista é selecionada. O especialista corrige os rótulos das amostras classificadas incorretamente (Linha 7). As amostras selecionadas são incluídas ao conjunto de treinamento \mathcal{L}_1 (Linha 8). O ciclo se repete até que o especialista esteja satisfeito com o desempenho do classificador (Linhas 4-9).

5.5 EXPERIMENTOS E RESULTADOS

Foram realizados experimentos com diferentes estratégias de aprendizado ativo e classificadores, utilizando os conjuntos de dados de folhas de plantas (LEA-53) e de enteroparasitos (PAR-16, PAR-15, EGG-9, EGG-8, PRO-7, PRO-6 e LAR-2).

Para todos os experimentos foram considerados 4 classificadores: SVM, k -NN, RF e NB e obtida a média das acurácias de 5 execuções para cada classificador. Os conjuntos de treinamento e de teste foram divididos de maneira estratificada na proporção 80% e 20%, respectivamente.

Algoritmo 6: Estratégia de Seleção - MST-CLU-RDS

Entrada: conjunto de treinamento organizado \mathcal{F} e conjunto de raízes \mathcal{R} .

Saída: classificador treinado.

Dados: conjunto de treinamento anotado \mathcal{L}_1 e conjunto de amostras selecionado \mathcal{L}'_1 .

```

1 início
2    $\mathcal{L}_1 \leftarrow \mathcal{R}$ ;
3   Especialista anota as amostras em  $\mathcal{L}_1$ ;
4   repita
5     (Re)-treinar o classificador com  $\mathcal{L}_1$ ;
6      $\mathcal{L}'_1 \leftarrow$  novas amostras classificadas em classes distintas seguindo a ordem
       da lista pré-organizada  $\mathcal{F}$ ;
7     Especialista corrige os rótulos das amostras em  $\mathcal{L}'_1$ ;
8      $\mathcal{L}_1 \leftarrow \mathcal{L}_1 \cup \mathcal{L}'_1$ ;
9   até que especialista satisfeito;
10 fim
  
```

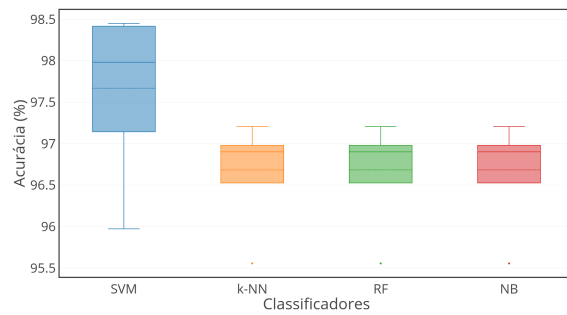
A Figura 28 ilustra as médias das acurácias sem o uso de estratégias de aprendizado ativo, obtidas com os classificadores SVM, k -NN, RF e NB, nos conjuntos de folhas de plantas (LEA-53) e de enteroparasitos (PAR-16, PAR-15, EGG-9, EGG-8, PRO-7, PRO-6 e LAR-2).

Dentre os classificadores analisados, o SVM obteve os melhores valores de acurácia. Portanto, tal classificador será considerado para os demais experimentos envolvendo as estratégias de aprendizado ativo apresentados nessa seção.

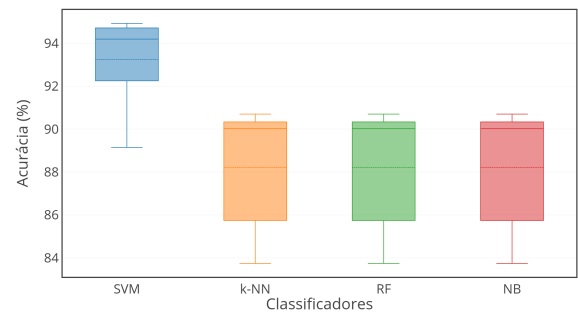
A técnica de *clustering* considerada para todos os experimentos envolvendo técnicas de aprendizado ativo que realizam o agrupamento dos dados como etapa inicial foi o *k-means*. O número de *clusters* (k) adotado nos experimentos foi definido como sendo o dobro do número de classes existentes em cada conjunto analisado, de forma a obter amostras representativas de cada classe na primeira iteração de aprendizado.

Diferentes estratégias de aprendizado ativo foram avaliadas, incluindo as novas propostas: MST-CLU-DS (Seção 5.3), MST-CLU-DDE (Seção 5.2) e MST-CLU-RDS (Seção 5.4); as técnicas clássicas de medida de incerteza: EN, MS e LC (Subseção 2.2.1); as novas variações das técnicas clássicas: EN-CLU, MS-CLU e LC-CLU (Seção 5.1), outras técnicas da literatura: RBE (Subseção 2.2.3), DBE (Subseção 2.2.4), MST-BE (Subseção 2.2.5) e RDS (Subseção 2.2.6).

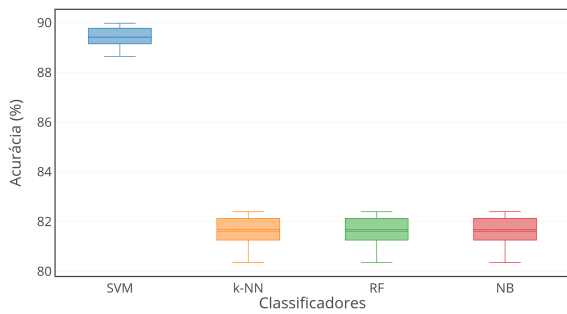
Os resultados obtidos pelas técnicas RBE, DBE e MST-BE não serão apresentados nessa seção, pois utilizam apenas parte reduzida dos conjuntos (apenas amostras de fronteira). Como consequência, amostras importantes podem ser descartadas do processo de aprendizado, impactando negativamente no desempenho do classificador (SAITO et al., 2015).



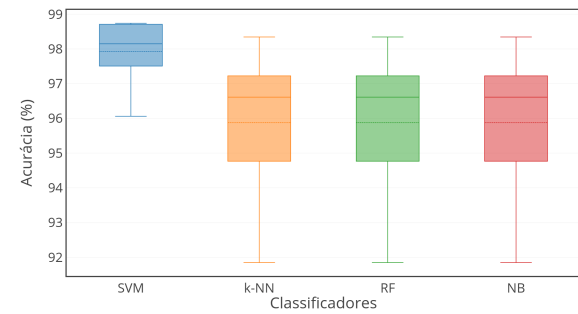
(a) LEA-53



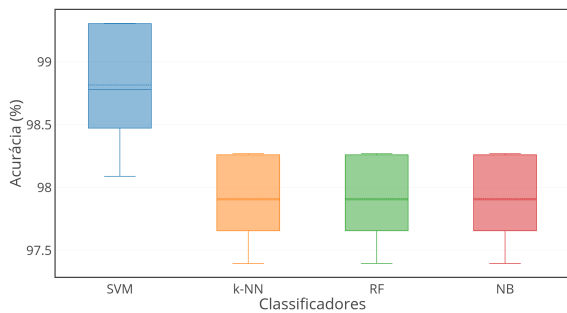
(b) PAR-16



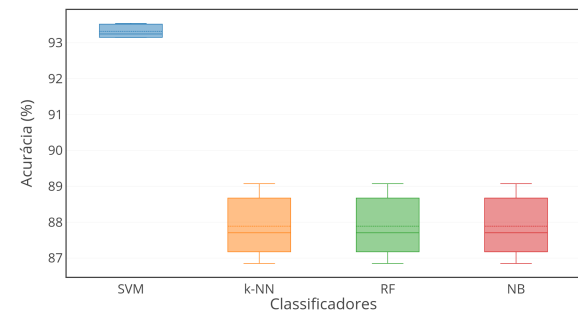
(c) PAR-15



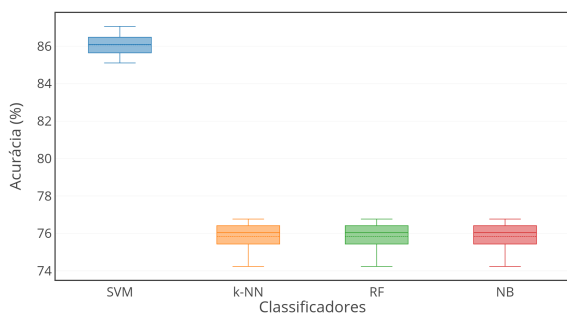
(d) EGG-9



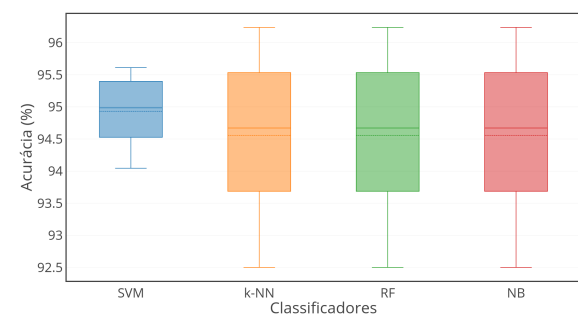
(e) EGG-8



(f) PRO-7



(g) PRO-6



(h) LAR-2

Figura 28: Médias das acurácias obtidas sem o uso de estratégias de aprendizado ativo com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

Acurácia, tempo de organização, tempo de seleção e tempo de classificação foram as métricas consideradas para as análises e comparações das estratégias de aprendizado ativo. As Figuras 29-31 comparam as médias das acurácias obtidas pelas estratégias clássicas de apren-

dizado ativo (EN, MS, LC) e pelas respectivas variações propostas (EN-CLU, MS-CLU e LC-CLU).

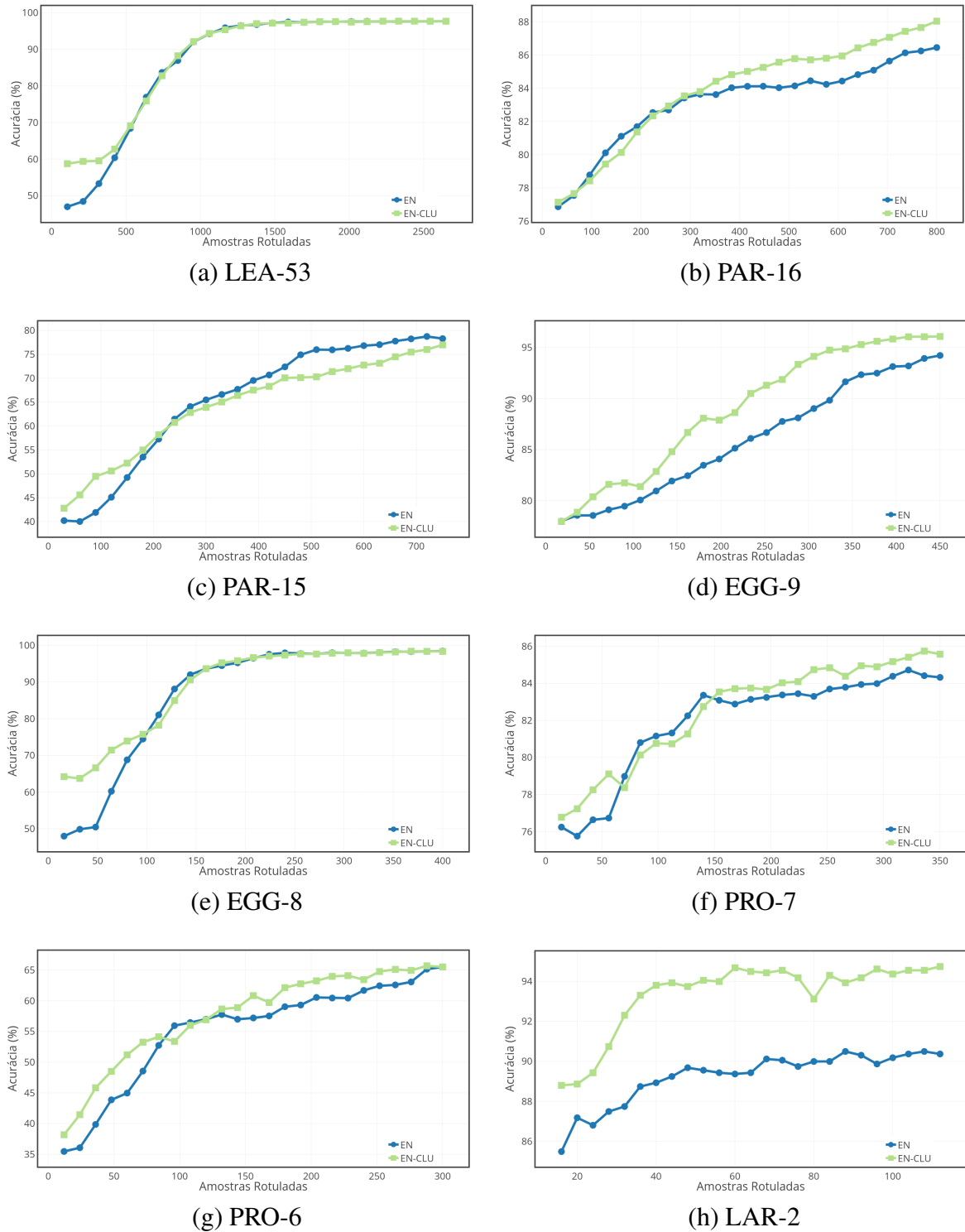


Figura 29: Médias das acurácias obtidas pelas estratégias de aprendizado ativo EN e EN-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

Os resultados obtidos pelas propostas de variações das técnicas clássicas (EN-CLU,

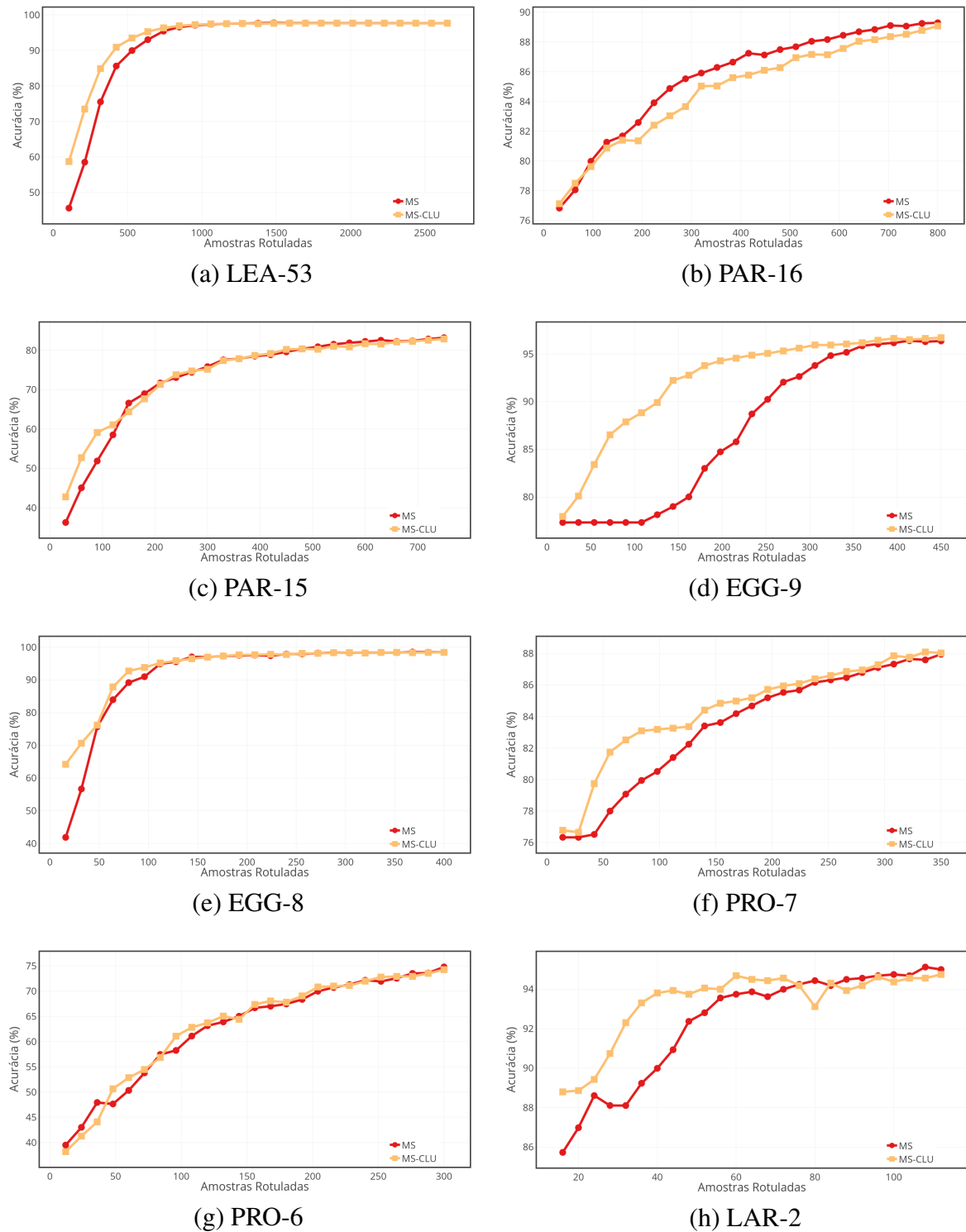


Figura 30: Médias das acurácias obtidas pelas estratégias de aprendizado ativo MS e MS-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

LC-CLU, MS-CLU), na maioria dos cenários analisados, exibem ganhos consideráveis na iteração inicial do aprendizado quando comparados com suas respectivas implementações originais. A estratégia de selecionar as raízes dos *clusters* mostra-se mais robusta. Ao explorar

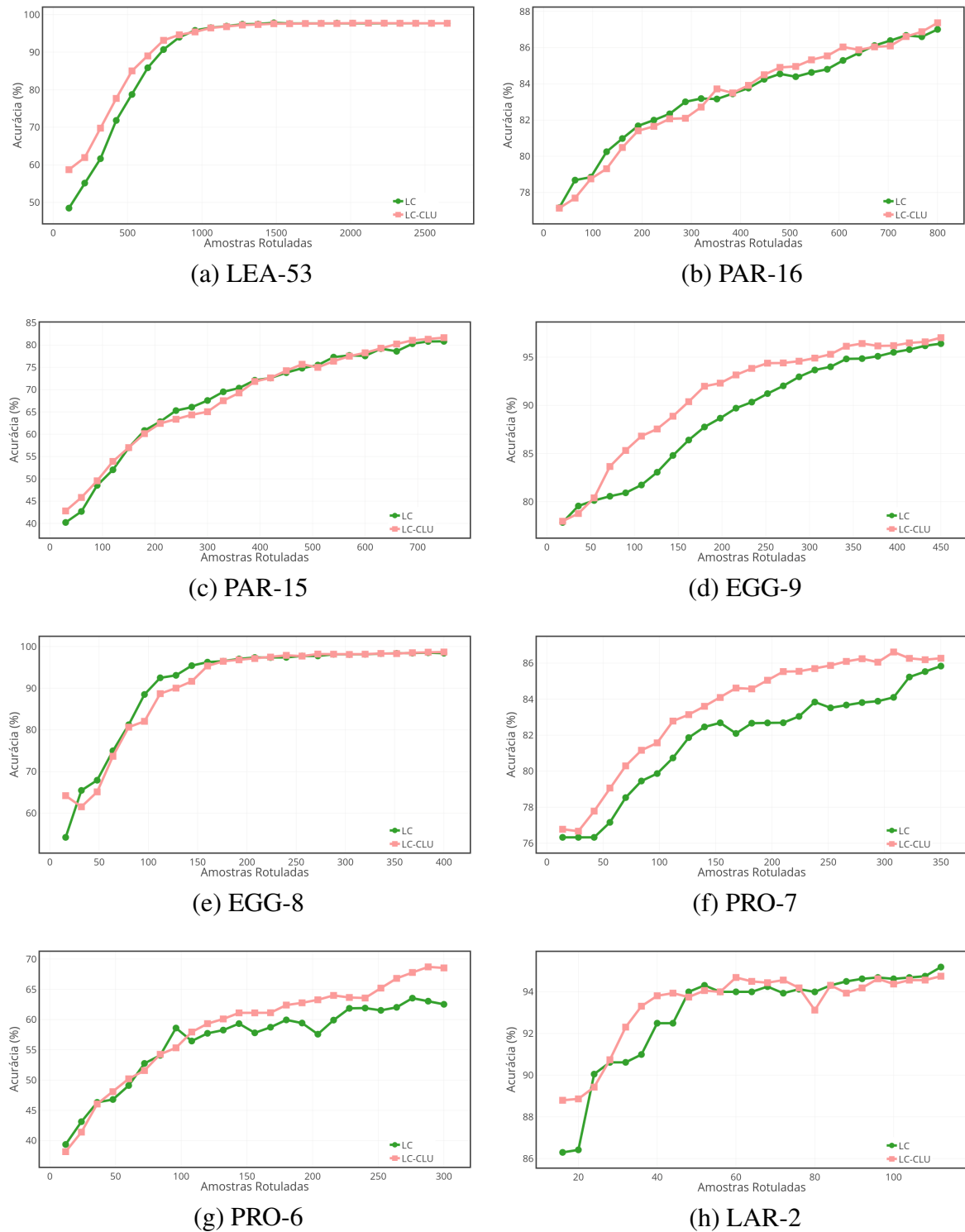


Figura 31: Médias das acurácias obtidas pelas estratégias de aprendizado ativo LC e LC-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

a diversidade obtém vantagem quando comparado com a inicialização aleatória proposta pelos métodos originais (EN, LC e MS).

Dando continuidade nas análises, foi avaliado quais dentre as propostas de variações das técnicas clássicas obtêm os melhores valores de acurácias. A Figura 32 ilustra as médias das acurácias obtidas pelas estratégias de aprendizado ativo EN-CLU, MS-CLU e LC-CLU utilizando o classificador SVM nos conjuntos de folhas de plantas (LEA-53) e de enteroparasitos (PAR-16, PAR-15, EGG-9, EGG-8, PRO-7, PRO-6 e LAR-2). É possível observar que a técnica MS-CLU, de forma geral, destaca-se das demais obtendo acurácias elevadas mais rapidamente (com menos iterações).

Dentre as estratégias de aprendizado ativo (RBE, DBE, MST-BE e RDS) da literatura baseadas em agrupamento e selecionadas para esse trabalho, a estratégia RDS obteve, de forma geral, os melhores valores de acurácias. Portanto, a mesma será apresentada nos comparativos com as novas propostas MST-CLU-DS, MST-CLU-DDE e MST-CLU-RDS, bem como com a estratégia MS-CLU (a qual também apresentou os melhores resultados, conforme Figura 32). A Figura 33 ilustra as médias das acurácias obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS utilizando o classificador SVM nos conjuntos de folhas de plantas (LEA-53) e de enteroparasitos (PAR-16, PAR-15, EGG-9, EGG-8, PRO-7, PRO-6 e LAR-2). Os demais resultados envolvendo as estratégias de aprendizado ativo (RBE, DBE, MST-BE e MS-CLU) são apresentados no Apêndice A.

Para o conjunto de folhas de plantas (LEA-53) e, de forma geral, para os conjuntos de enteroparasitos com a inclusão de impurezas (PAR-16, EGG-9, PRO-7), os métodos MS-CLU e RDS apresentaram melhor desempenho (Figuras 33b, 33d e 33f). No entanto, para o conjunto LAR-2 (Figura 33h) o método proposto MST-CLU-DDE, juntamente com MS-CLU, apresentaram resultados superiores ao RDS. Para os demais conjuntos de enteroparasitos (sem impurezas) as técnicas propostas apresentam maiores acurácias mais rapidamente. Por exemplo, a Figura 33g ilustra o desempenho superior da estratégia MST-CLU-RDS, em que apresenta acurácia superior com menos iterações de aprendizado (menos amostras rotuladas), enquanto as demais estratégias de aprendizado requerem mais iterações para atingir valores de acurácias similares. Também é possível observar que na medida em que as iterações avançam as estratégias vão convergindo para acurácias similares.

Além da acurácia, outras métricas também foram avaliadas. A Figura 34 apresenta a síntese dos resultados considerando as métricas de acurácia e de tempos computacionais de organização, seleção e classificação ao longo das iterações, utilizando o classificador SVM e as técnicas de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS. Os resultados apresentados foram normalizados e representam as médias de cada métrica. Maiores distâncias em relação ao centro da circunferência indicam os melhores resultados.

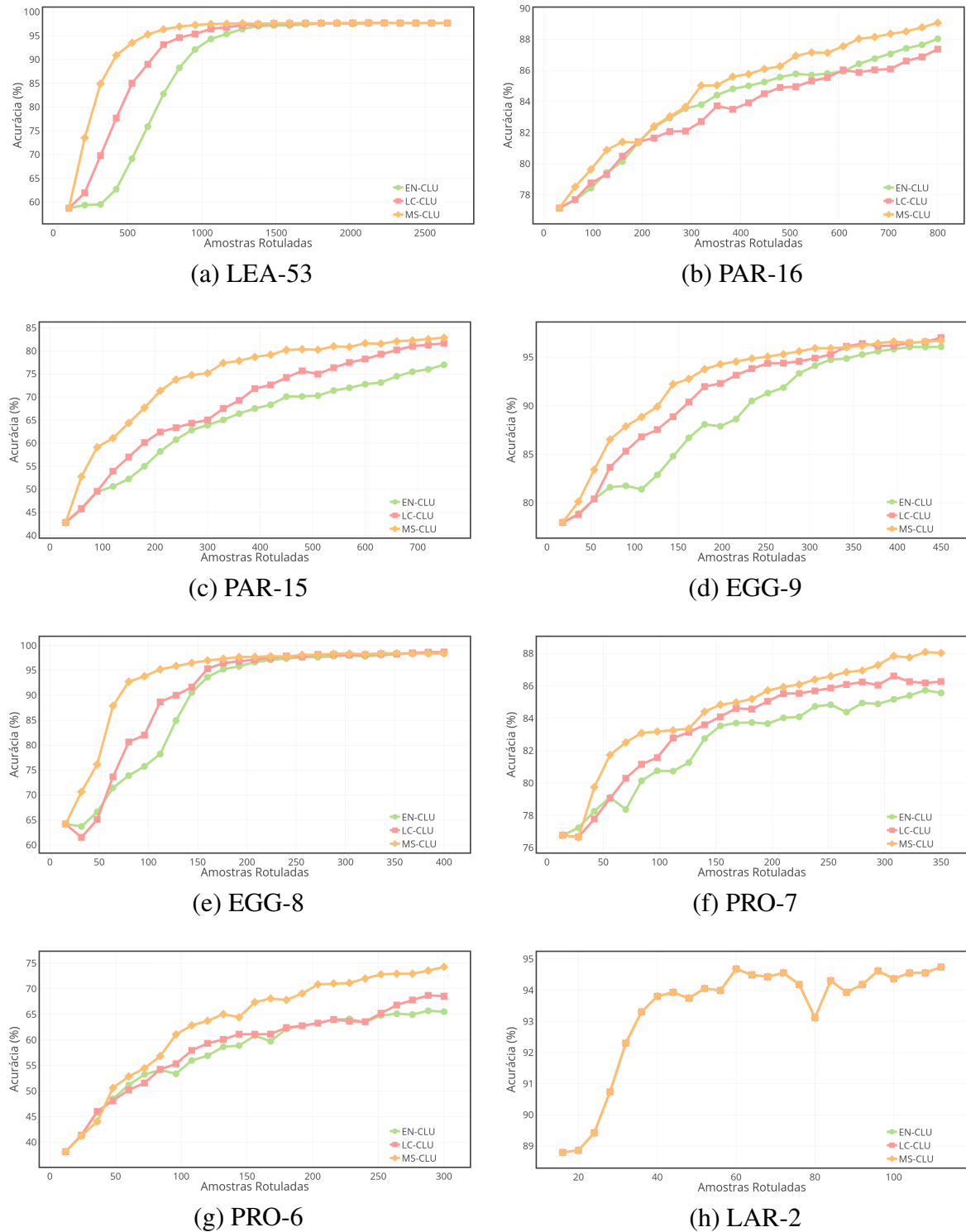


Figura 32: Médias das acurácias obtidas pelas estratégias de aprendizado ativo EN-CLU, MS-CLU e LC-CLU com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

Considerando o tempo de organização, as estratégias propostas (MST-CLU-DDE, MST-CLU-DS e MST-CLU-RDS) apresentam resultados inferiores aos das estratégias RDS e MS-CLU. Esse custo maior de tempo de organização ocorre pois as estratégias propostas reque-

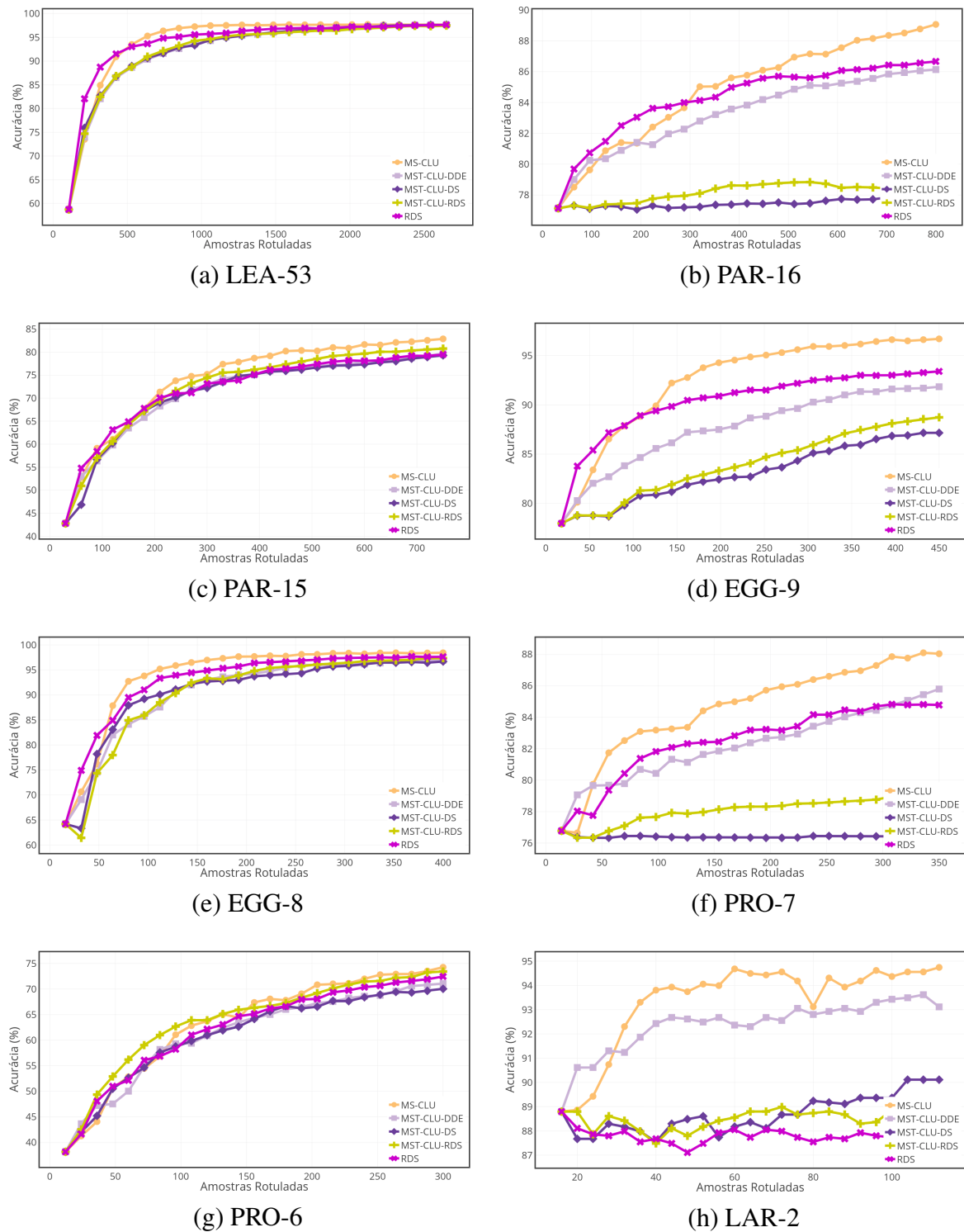


Figura 33: Médias das acurácias obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

rem que o algoritmo da MST seja aplicado múltiplas vezes. A quantidade é determinada pelo número de *clusters* definido na inicialização do método.

Em termos de tempo de seleção, a estratégia MST-CLU-DS, ao não consultar o classificador para tomada de decisão de seleção das amostras, é consideravelmente superior em todos os conjuntos analisados. Além disso, apesar de consultar o classificador para seleção das amostras, MST-CLU-DDE também apresenta desempenho superior em relação às estratégias MS-CLU, RDS e MST-CLU-RDS. A estratégia MS-CLU apresenta desvantagem e os piores resultados de tempo de seleção, pois requer que todas as amostras sejam avaliadas pelo classificador a cada iteração do aprendizado.

O tempo de classificação obtido por cada estratégia nos conjuntos analisados não apresenta um padrão bem delimitado. A provável causa desse comportamento pode estar associada a detalhes de implementação do classificador. Outro fator que dificulta a análise dessa métrica é a escala da medida de tempo (muito pequena). Os valores em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias (MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS) encontram-se no Apêndice I.

Resultados considerando as métricas de acurácia e de tempos computacionais (de organização, seleção e classificação) utilizando outros classificadores (k -NN, RF, NB) podem ser observados no Apêndice (Apêndices B-D). Além disso, o desempenho de cada estratégia de aprendizado ativo ao longo das iterações de aprendizado foi avaliado para todos os conjuntos de dados considerando diferentes classificadores (SVM, k -NN, RF, e NB). A Figura 35 apresenta as acurácias obtidas pela estratégia MS-CLU utilizando tais classificadores em cada conjunto de dados. Analisando os gráficos da Figura 35 pode-se observar que o classificador SVM apresenta, de forma geral, maiores valores de acurácia. O mesmo pode ser observado para as demais estratégias de aprendizado ativo analisadas (Apêndices E-H).

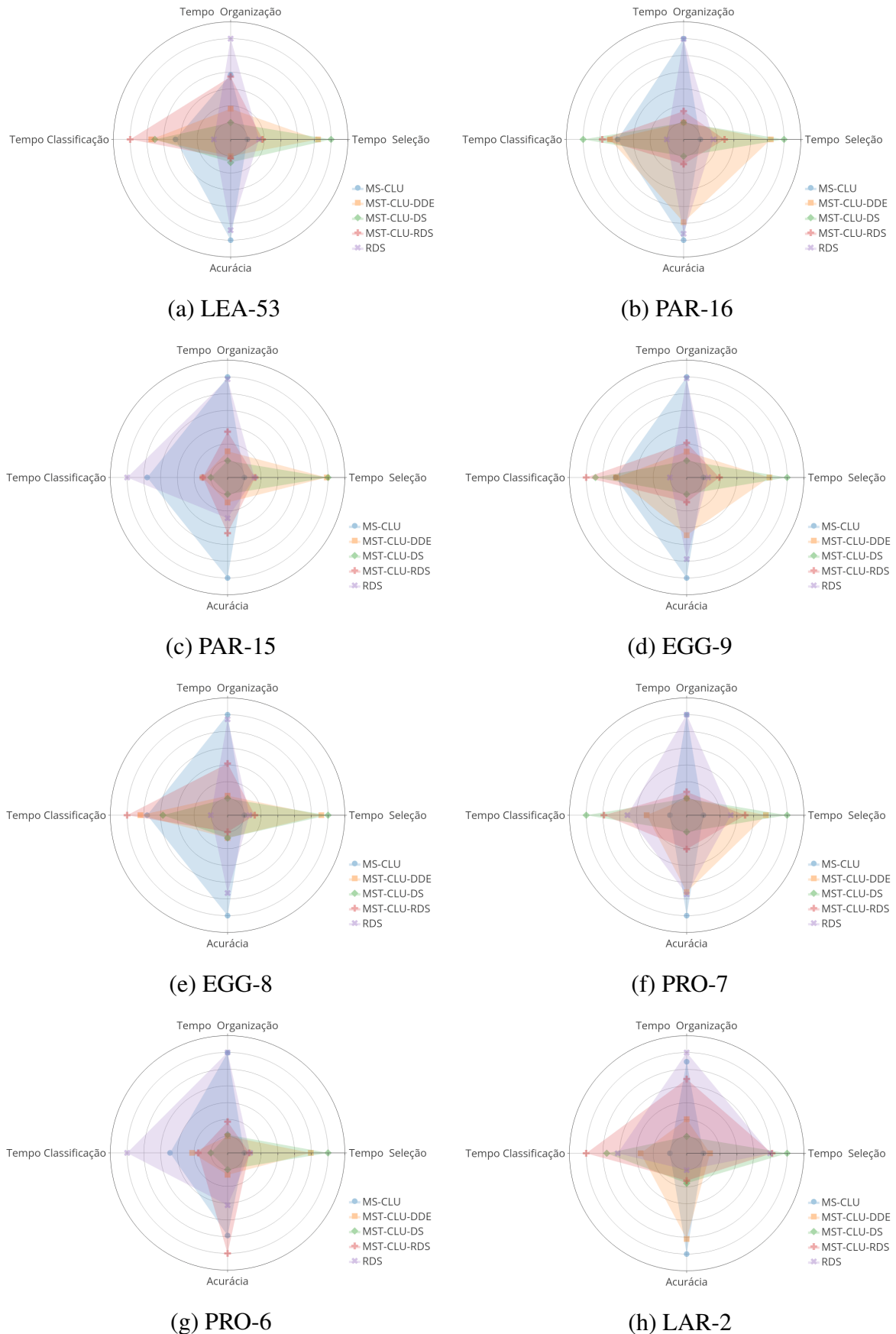
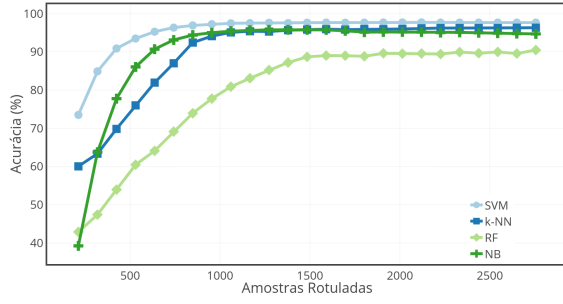
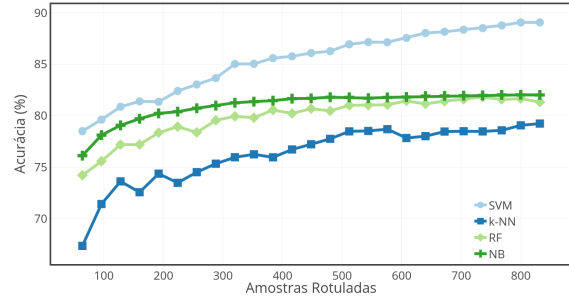


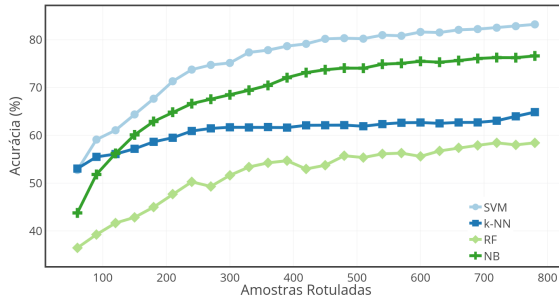
Figura 34: Médias de acurácias e de tempos de organização, seleção e classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.



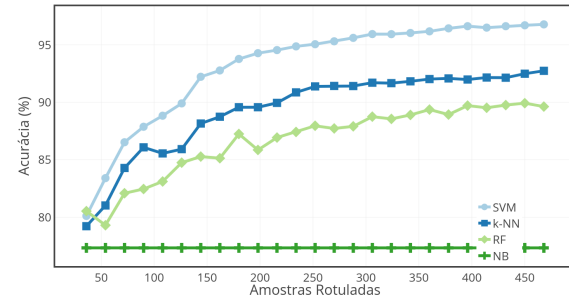
(a) LEA-53



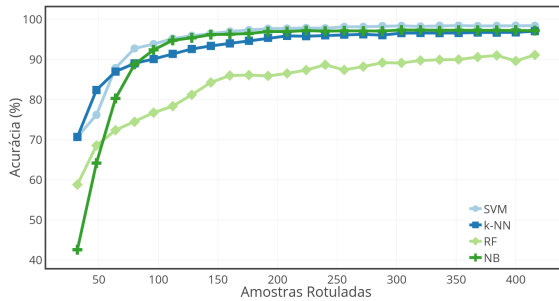
(b) PAR-16



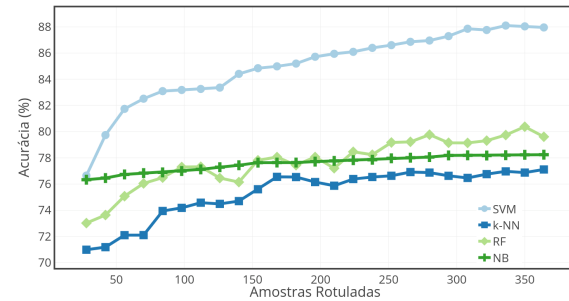
(c) PAR-15



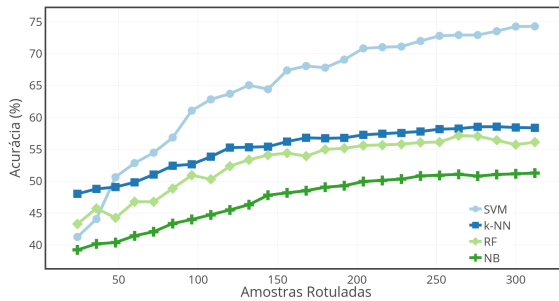
(d) EGG-9



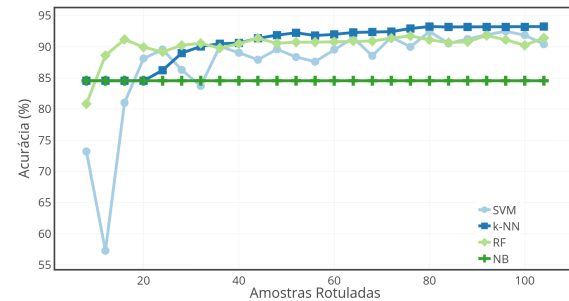
(e) EGG-8



(f) PRO-7



(g) PRO-6



(h) LAR-2

Figura 35: Médias das acurácias obtidas pela estratégia de aprendizado ativo MS-CLU com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

6 CONTRIBUIÇÃO 3 - FRAMEWORK PARA SÍNTESE E CLASSIFICAÇÃO DE ENTEROPARASITOS

6.1 METODOLOGIA PROPOSTA

Uma nova abordagem para classificação de imagens foi desenvolvida, explorando estratégias de aprendizado ativo, de forma a selecionar amostras mais informativas, nos processos de sintetização de imagens e de aprendizados tradicional e em profundidade (ALVES et al., 2019c).

Para lidar com os problemas mencionados anteriormente, relacionados à obtenção de acurácias baixas, dada a escassez de imagens informativas utilizadas no aprendizado do classificador, foi proposta a criação de um sintetizador de imagens robusto, capaz de gerar e fornecer, de maneira não supervisionada, conjuntos de imagens sintéticas mais informativas a um especialista. Sendo assim, após anotadas, as mesmas são utilizadas para o treinamento do classificador.

A Figura 36 ilustra os fluxos da abordagem de aprendizado proposta, destacando cada uma das etapas. Inicialmente, tem-se como entrada imagens originais não rotuladas presentes no conjunto de dados inicial (Etapa 1).

A partir das imagens originais, a Etapa 2 consiste na extração de características que melhor descrevam tais imagens. Diferentes extratores podem ser considerados nessa etapa, de acordo com o domínio do conjunto de imagens. Técnicas de *Transfer Learning* têm sido bem sucedidas e também são exploradas para obtenção das características, conforme resultados apresentados no Capítulo 4.

Na Etapa 3, uma dada técnica de aprendizado não supervisionado deve ser aplicada ao conjunto de dados inicial, de forma a melhorar o processo de geração de novas imagens sintéticas. Realizar o agrupamento das imagens possibilita a geração de imagens de cada um dos grupos, aumentando a probabilidade de gerar imagens de cada uma das classes existentes no conjunto. Nessa etapa qualquer técnica de agrupamento pode ser aplicada, dentre elas: *k-means* (HARTIGAN; WONG, 1979), *Optimum-Path Forest (OPF)* (ROCHA et al., 2009), *Self-Organizing Map (SOM)* (KOHONEN, 1998), entre outras.

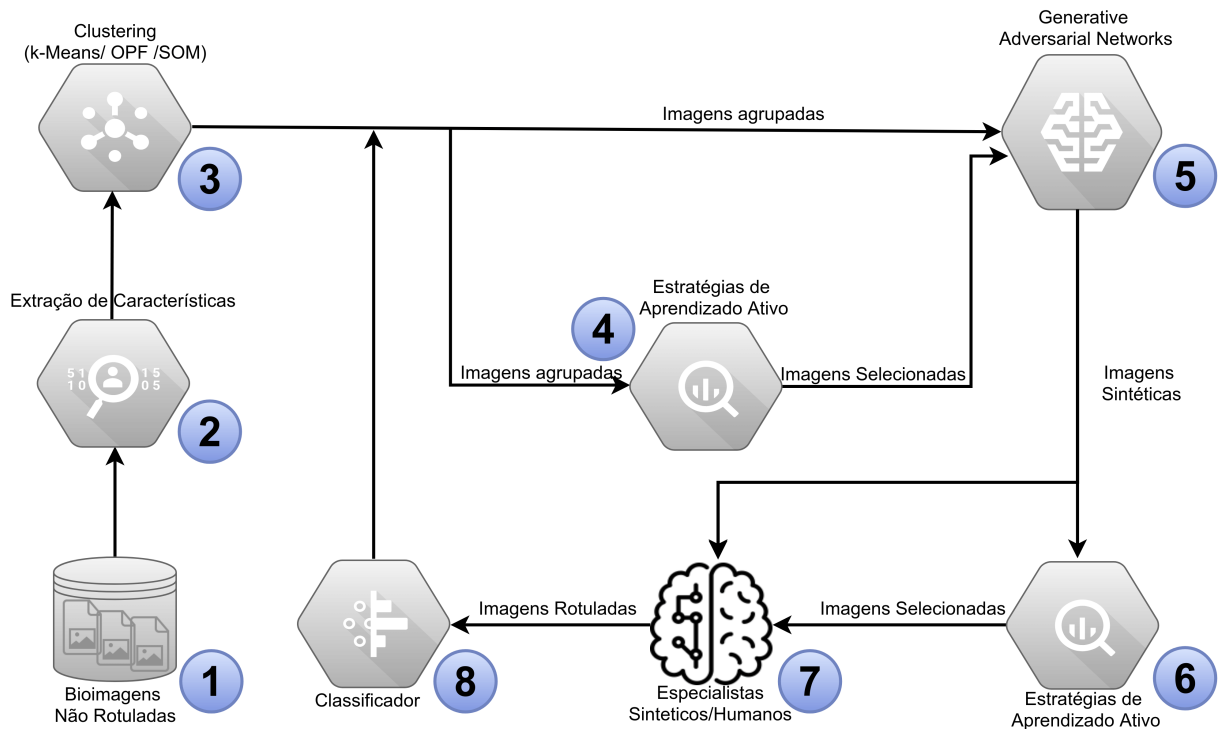


Figura 36: Pipeline da abordagem de aprendizado proposta para anotação de imagens, envolvendo a aplicação de estratégias de aprendizado ativo na construção de bases de imagens sintéticas e no aprendizado de classificadores.

Após realizar o agrupamento das imagens, estratégias de aprendizado ativo podem ser exploradas (Etapa 4), para refinar e selecionar as imagens mais informativas a serem utilizadas como base para geração das novas imagens sintéticas. Trata-se de uma etapa opcional, na qual uma avaliação experimental pode ser realizada de forma a analisar o impacto e as diferenças no desempenho. É então possível verificar se a utilização de todas as amostras do conjunto é prejudicial, em termos de eficiência dada a quantidade de imagens a serem processadas, bem como em eficácia dada a utilização de imagens desnecessárias que podem impactar de forma negativa o desempenho do classificador. É possível também, avaliar se melhorias podem ser obtidas por meio da seleção de imagens mais representativas (de cada classe) e informativas (de difícil classificação) no processo de geração de um novo conjunto de imagens sintéticas.

A partir das imagens agrupadas e/ou selecionadas (Etapa 3 e/ou Etapa 4), *Generative Adversarial Networks* - GANs (descritas na Seção 2.3.2) devem ser utilizadas (Etapa 5) para a geração das imagens sintéticas. A abordagem envolvendo estratégias de aprendizado ativo juntamente com GANs foi utilizada, e de certa forma bem sucedida, em (ZHU; BENTO, 2017). No entanto, a abordagem proposta pelos autores apresenta restrições, como condução de experimentos aplicados apenas no contexto de classificação binária de imagens, e baixa correlação entre as imagens de classes distintas. Além disso, estudos ou outros métodos envolvendo GANs

não foram explorados.

Para prover melhorias na qualidade das imagens geradas, o presente trabalho visa explorar novos conceitos de GANs, como *Wasserstein Generative Adversarial Networks* (WGAN) (GULRAJANI et al., 2017), *Wasserstein Generative Adversarial Networks with Gradient Penalty* (WGAN-GP) (GULRAJANI et al., 2017), *Boundary Equilibrium Generative Adversarial Networks* (BEGAN) (BERTHELOT et al., 2017), que têm se destacado na geração de imagens mais realistas e sem borrões. Além das variantes das GANs e as diferentes configurações (hiperparâmetros), estratégias de aprendizado ativo mais eficientes e efetivas são propostas (vide Capítulo 5) e utilizadas, diferentemente das apresentadas em (ZHU; BENTO, 2017).

Após a criação das imagens sintéticas por meio das GANs, opcionalmente a Etapa 6 pode ser realizada, na qual estratégias de aprendizado ativo podem também ser exploradas, de forma a selecionar as imagens sintéticas mais informativas a serem exibidas ao especialista. O especialista, por sua vez, realizará uma verificação (Etapa 7), anotando as imagens na primeira iteração, e corrigindo ou confirmando os rótulos nas iterações seguintes. Da mesma forma, conforme mencionado na Etapa 4, uma avaliação experimental pode ser realizada de forma a verificar a existência ou não de benefícios com a inclusão de estratégias de aprendizado ativo para seleção das amostras sintéticas na Etapa 6.

As imagens devidamente rotuladas, após verificação do especialista na Etapa 7, são incorporadas ao conjunto de treinamento e gera-se uma instância do classificador (Etapa 8). O processo de aprendizado é repetido até que a condição de parada seja satisfeita. Por exemplo, até que o classificador obtido apresente desempenho esperado e satisfatório, de acordo com o domínio de aplicação ou que seja obtido o conjunto de imagens sintéticas desejado.

A abordagem de aprendizado proposta, consiste em explorar o uso conjunto de estratégias de aprendizado ativo e aprendizado profundo para melhorar os processos de síntese de imagens e de aprendizado do classificador e, conseqüentemente, o processo de anotação das imagens.

6.2 ANÁLISE DAS MÉTRICAS DE AVALIAÇÃO DAS GANs

Foram realizados experimentos envolvendo *data augmentation* com o uso das GANs nos conjuntos de dados propostos. O intuito inicialmente foi verificar se a qualidade visual das imagens geradas seria adequada, no contexto de aprendizado ativo, com a interação de um especialista anotando as mesmas para treinamento do classificador durante o ciclo de aprendizado.

No entanto, foram encontrados vários desafios, dentre eles a dificuldade no treina-

mento das GANs, no qual nem sempre foi possível estabilizá-lo, ou seja, a rede geradora estagna e gera sempre o mesmo padrão. Essas imagens geradas, normalmente são apenas ruídos e não possuem qualquer relação visual com as imagens presentes no conjunto de dados original.

Outra dificuldade refere-se à alta sensibilidade em ajustes nos hiperparâmetros das GANs. Pequenas oscilações, por exemplo, no tamanho e escala do vetor de ruído da entrada da rede geradora, geram resultados completamente imprevisíveis.

Algumas arquiteturas consideradas apresentaram melhores resultados com uma qualidade visual das imagens geradas, minimamente semelhantes às presentes no conjunto de dados original. Foram analisadas 4 arquiteturas de GANs: GAN (GOODFELLOW et al., 2014), DCGAN (RADFORD et al., 2015), WGAN (GULRAJANI et al., 2017) e BEGAN (BERTHELOT et al., 2017). A arquitetura que apresentou os melhores resultados na qualidade visual das imagens geradas foi a BEGAN. A Figura 38 apresenta exemplos das imagens reais e sintéticas (geradas pela arquitetura BEGAN) para cada uma das classes de enteroparasitos.

Outro desafio no cenário das GANs, por exemplo, está na escassez de métricas que avaliam a relação da qualidade visual das imagens geradas com o desempenho das mesmas em uma tarefa de classificação. Ainda não existe um consenso e uma métrica de qualidade amplamente utilizada, por exemplo, no contexto de imagens.

Em Salimans et al. (2016) os autores propõem a métrica *Inception Score* que utiliza a rede neural Inception pré-treinada no conjunto de dados Imagenet para atribuir uma pontuação para a imagem. No entanto, alguns trabalhos (HEUSEL et al., 2017; BARRATT; SHARMA, 2018) apontam que essa métrica não captura a similaridade entre as imagens geradas e as imagens reais e não fornece orientação útil ao comparar variações das GANs. A maioria dos trabalhos tentam justificar as métricas propostas mostrando uma forte correlação com a avaliação humana (SALIMANS et al., 2016; LOPEZ-PAZ; OQUAB, 2016). Contudo, a avaliação humana tende a ser tendenciosa em relação à qualidade visual das amostras geradas e negligencia características gerais de distribuição, as quais são importantes para o aprendizado. Portanto, embora algumas métricas demonstrem ser eficazes em várias tarefas, não se sabe os cenários exatos em que suas pontuações são significativas, podendo gerar interpretações equivocadas.

Nesse trabalho, foram avaliadas algumas das principais métricas para análise da qualidade e diversidade de imagens sintéticas geradas pelas GANs. A Figura 37 ilustra o fluxo proposto para análise dessas métricas. Inicialmente, tem-se a geração das imagens sintéticas a partir das imagens reais, seguida da extração e obtenção das características (por meio da arquitetura Inception-v3 e PCA) e por fim o cálculo das métricas.

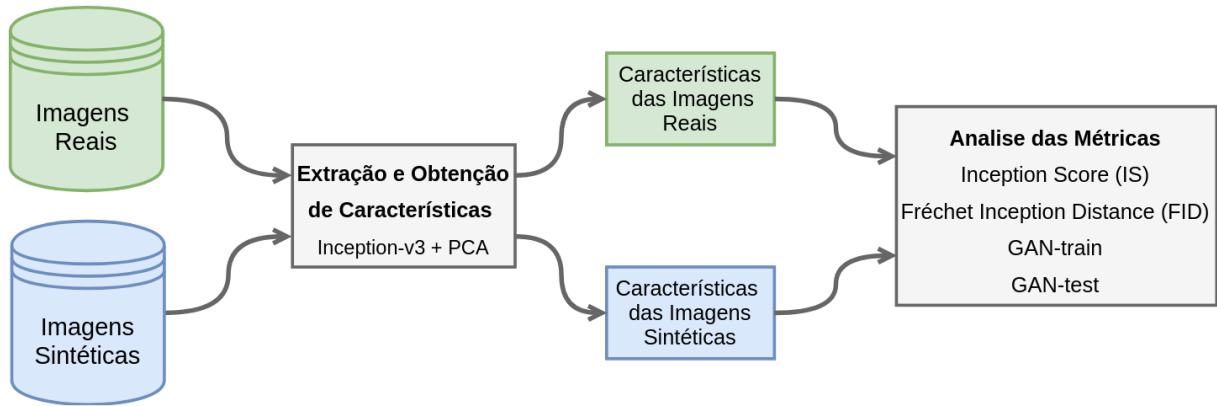


Figura 37: Fluxo para análise das métricas utilizadas para avaliação das imagens geradas pelas GANs.

Tabela 9: Métricas *Inception Score (IS)*, *Mode Score (MD)* e *Fréchet Inception Distance (FID)* para avaliação das imagens (de cada classe) de enteroparasitos geradas pela arquitetura BEGAN. São destacados os melhores (negrito) e os piores (underline) resultados obtidos por cada métrica.

Classe (PAR-15)	IS	MD	FID
S.stercoralis	1.435	0.881	0.191
H.diminuta	1.136	0.902	0.091
E.coli	1.155	0.978	0.097
E.nana	1.181	0.975	0.136
H.nana	1.215	0.891	0.130
A.lumbricoides	1.188	1.005	0.090
E.histolytica	<u>1.094</u>	1.065	0.062
B.hominis	1.127	0.943	0.085
Ancilostomideo	1.105	0.754	0.147
T.trichiura	1.143	0.968	0.071
E.vermicularis	1.237	<u>0.568</u>	0.165
S.mansoni	1.126	0.636	0.141
Giardia	1.222	0.871	<u>0.225</u>
Taenia	1.188	1.071	0.094
I.butshlii	1.188	1.084	0.072

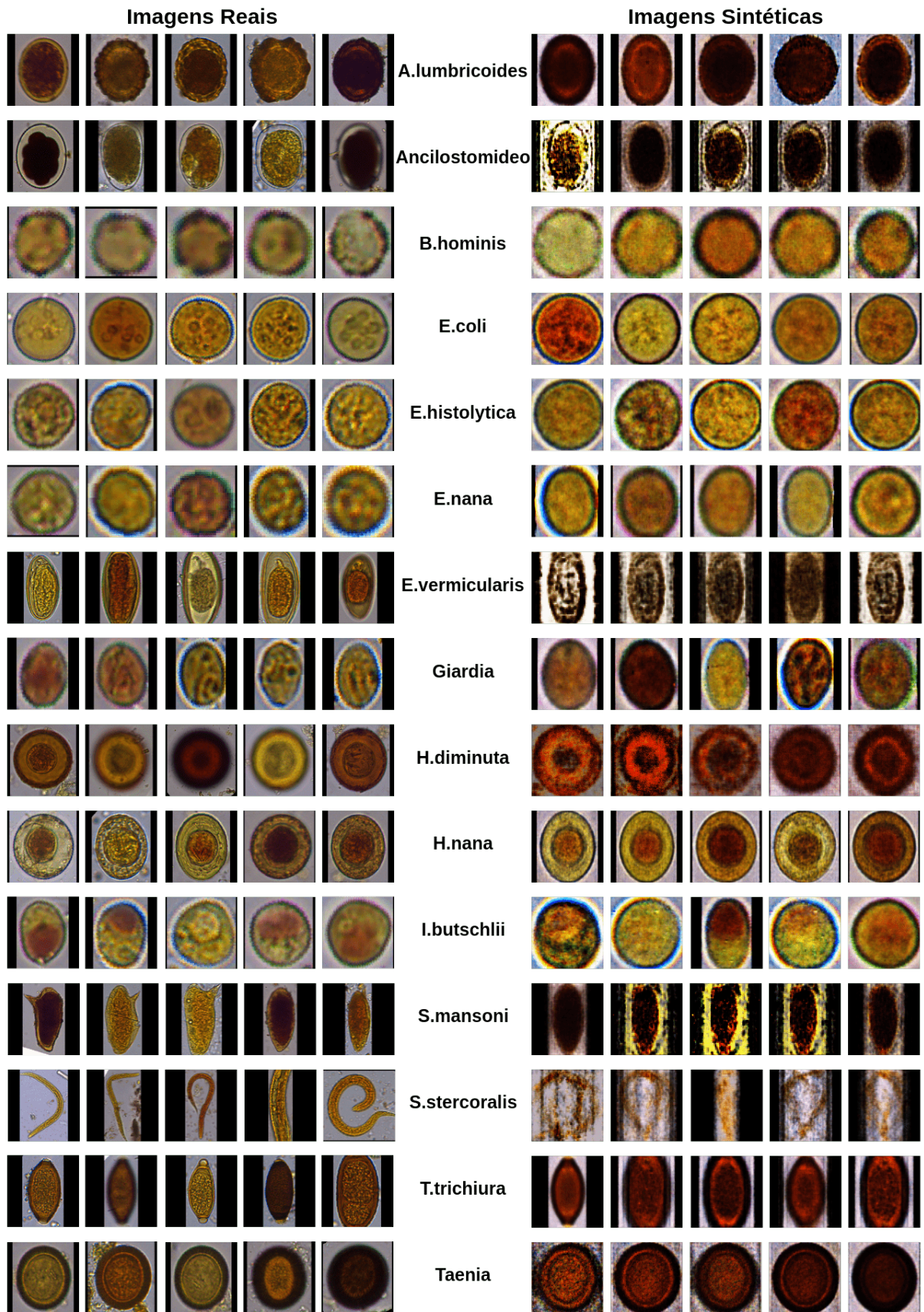


Figura 38: Exemplos de imagens (reais e sintéticas) do conjunto de enteroparasitos. As cinco primeiras imagens referem-se às imagens reais e as cinco imagens seguintes correspondem às imagens sintéticas geradas pela BEGAN.

A Tabela 9 resume os resultados das métricas *Inception Score* (IS), *Mode Score* (MD) e *Fréchet Inception Distance* (FID) para avaliação das imagens sintéticas de (cada classe de) enteroparasitos geradas pela arquitetura BEGAN. Para essa análise foram consideradas 1.000 imagens sintéticas para cada classe. Valores maiores para IS e MD representam melhor desempenho e valores menores para FID representam melhor desempenho, segundo a avaliação dessas métricas. São destacados os melhores (negrito) e os piores (underline) resultados obtidos por cada métrica.

Pode-se observar (Tabela 9) que as imagens sintéticas da classe **S.stercoralis** foram consideradas as melhores segundo a métrica IS. Já a métrica MD, apontou a classe **I.butchlii** como sendo a melhor classe de imagens sintéticas. No entanto, a terceira métrica avaliada (FID) também divergiu das demais ao apresentar a classe **E.histolytica** como sendo a classe que apresenta maior qualidade e diversidade, dentre as classes de imagens geradas.

As métricas IS, MD e FID, apesar de populares e bastante utilizadas na literatura, não se mostram adequadas no contexto de imagens sintéticas quando aplicadas no domínio de enteroparasitos, visto que há uma discordância em apontar qual classe de imagens sintéticas apresenta melhor qualidade e diversidade. Portanto, as demais análises consideram as métricas *GAN-train* e *GAN-test* (SHMELKOV et al., 2018), que se mostram mais resilientes e adequadas para o objetivo final da abordagem proposta, o qual consiste em obter um classificador de padrões mais robusto.

A Tabela 10 resume os resultados de acurácias obtidas pelo classificador SVM, para cada conjunto de enteroparasitos LAR-2, EGG-8, EGG-9, PRO-6, PRO-7, PAR-15 e PAR-16, considerando as abordagens *baseline* (contendo apenas amostras reais) e as arquiteturas GAN, DCGAN, WGAN, WGAN-GP e BEGAN (utilizando as imagens sintéticas e as métricas *GAN-train* e *GAN-test*). Os resultados da abordagem *baseline* referem-se à média de 5 execuções para cada conjunto, considerando a separação estratificada de 80% das amostras reais para treinamento e 20% para testes. Para os resultados com a métrica *GAN-train*, é considerado o treinamento com 1.000 imagens sintéticas de cada classe presente no conjunto avaliado. Por exemplo, o conjunto EGG-8, possui 8 classes distintas. Portanto, foram utilizadas 1.000 amostras sintéticas de cada classe totalizando 8.000 amostras de treinamento, e a acurácia final foi obtida com o conjunto de teste contendo 20% das amostras de imagens reais. Os resultados considerando a métrica *GAN-test* foram obtidos a partir do treinamento com o conjunto de imagens reais e testes com as imagens sintéticas.

A arquitetura BEGAN, de forma geral, apresentou os melhores resultados considerando as métricas e conjuntos analisados (Tabela 10). Por exemplo, para os conjuntos EGG-8,

Tabela 10: Acurácias (%) obtidas pelo classificador SVM, para cada conjunto de enteroparasitos (LAR-2, EGG-8, EGG-9, PRO-6, PRO-7, PAR-15 e PAR-16), considerando as abordagens *baseline* (contendo apenas amostras reais) e as arquiteturas GAN, DCGAN, WGAN, WGAN-GP e BEGAN (utilizando imagens sintéticas e as métricas *GAN-train* e *GAN-test*).

Conjunto	Baseline	Métrica	GAN	DCGAN	WGAN	WGAN-GP	BEGAN
LAR-2	94.93	GAN-train	9.54	11.82	11.57	13.96	15.46
		GAN-test	1.10	2.32	2.61	2.31	1.92
EGG-8	98.82	GAN-train	12.31	11.71	14.74	16.93	17.56
		GAN-test	10.40	11.52	9.88	11.04	11.73
EGG-9	97.93	GAN-train	2.74	3.71	3.52	4.53	3.98
		GAN-test	2.26	2.69	2.19	3.35	2.75
PRO-6	86.07	GAN-train	15.65	20.40	18.77	21.36	22.61
		GAN-test	11.45	14.08	17.93	19.56	21.05
PRO-7	93.32	GAN-train	5.44	4.67	5.26	5.54	5.35
		GAN-test	3.06	3.72	4.48	4.19	4.51
PAR-15	89.41	GAN-train	5.41	5.14	5.31	6.86	6.57
		GAN-test	4.96	5.48	5.23	5.75	6.23
PAR-16	93.24	GAN-train	3.10	4.68	5.19	5.23	5.43
		GAN-test	9.70	9.46	9.73	11.21	12.39

PRO-6 e PAR-16, a BEGAN obteve os melhores resultados considerando ambas as métricas GAN-train e GAN-test, em que as melhores acurácias (22.61% e 21.05%, respectivamente) foram obtidos no conjunto PRO-6. De fato, a qualidade visual das imagens analisando cada grupo (larvas, ovos e protozoários) de enteroparasitos, utilizando a arquitetura BEGAN, de forma geral, todas as classes do grupo de protozoários (vide Tabela 2, Capítulo 3) apresentam imagens sintéticas (visualmente) mais similares às imagens reais em comparação às imagens geradas a partir dos demais grupos (de ovos e larvas) do conjunto de enteroparasitos (Figura 38).

Comparando as métricas GAN-train e GAN-test (Tabela 10), observa-se que treinar utilizando o conjunto de imagens sintéticas e testar utilizando o conjunto de imagens reais (i.e. GAN-train) apresenta melhores resultados para todos os conjuntos de enteroparasitos analisados, com exceção do conjunto PAR-16. Tal conjunto apresenta resultados diferentes dos demais, pois os melhores valores ocorreram quando o treinamento foi realizado utilizando o conjunto de imagens reais e o teste utilizando o conjunto de imagens sintéticas (GAN-test). Isso indica que a métrica GAN-test é mais resiliente e adequada para conjuntos consideravelmente grandes e desbalanceados, como é o caso do PAR-16 que possui 51.665 amostras, dentre as quais aproximadamente 77% estão presentes em uma única classe (impurezas).

Além do escopo de imagens, análises também foram realizadas com o conjunto CIFAR-10. A Tabela 11 apresenta essas análises, considerando as arquiteturas de GANs: GAN, DCGAN, WGAN, WGAN-GP e BEGAN; e as métricas: IS, MD, FID, GAN-train e GAN-test.

Pode-se observar que é quase unânime o melhor desempenho obtido pela arquitetura BEGAN, a qual somente foi superada pela WGAN-GP, segundo a métrica *GAN-test*.

Tabela 11: Resultados obtidos utilizando imagens reais (baseline) e imagens sintéticas por meio das arquiteturas GAN, DCGAN, WGAN, WGAN-GP e BEGAN, considerando os valores das métricas IS, MD, FID, GAN-train e GAN-test no conjunto de dados CIFAR-10. GAN-train e GAN-test são medidas de acurácia com o classificador SVM.

Modelo	IS	MD	FID	GAN-train	GAN-test
Imagens Reais	11.33	11.67	9.4	92.8	-
GAN	5.43	5.98	56.5	63.5	51.3
DCGAN	6.69	7.12	42.5	67.5	58.2
WGAN	5.77	5.87	25.4	65.6	61.3
WGAN-GP	8.29	8.21	22.1	76.1	85.0
BEGAN	8.43	8.39	19.8	79.5	83.4

Grandes esforços têm sido despendidos, envolvendo a escolha de arquitetura, implementações, treinamento e ajustes de hiper-parâmetros das GANs. No entanto, nenhuma amostra sintética de enteroparasitos conseguiu agregar significativamente a qualidade visual necessária com o bom desempenho segundo as métricas avaliadas.

Apesar disso, os resultados obtidos são consistentes aos reportados em alguns trabalhos da literatura (XU et al., 2018; SHMELKOV et al., 2018). É notável que, mesmo para um conjunto de dados bem comportado (balanceado) e simples (distinção entre classes) como o CIFAR-10, agregar ao treinamento imagens sintéticas geradas pelas GANs não melhora os valores de acurácias quando comparado aos *baselines* (sem a inclusão das imagens sintéticas). Além disso, é possível afirmar que o domínio e o conjunto de enteroparasitos apresentam um nível de complexidade maior, visto que ele pode estar significativamente desbalanceado (principalmente nos casos que consideram a classe de impurezas) e a distinção entre classes exigir um especialista.

Explorar novas arquiteturas de GANs pode ser uma alternativa promissora. Novos trabalhos da literatura tem mostrado resultados impressionantes quanto à qualidade das imagens geradas (KARRAS et al., 2017; BROCK et al., 2018). No entanto, o treinamento dessas arquiteturas pode ser inviabilizado pelas elevadas exigências de poder computacional.

Sendo assim, considerando as discussões apresentadas, a arquitetura BEGAN foi selecionada para dar prosseguimento nas demais análises e experimentos para avaliação da abordagem proposta, pois, dentre as arquiteturas analisadas, foi a que obteve os melhores resultados segundo as métricas avaliadas.

Resultados positivos da abordagem proposta dependem não só da boa qualidade visual das imagens geradas, para que um especialista seja capaz de rotulá-las adequadamente, mas

também da representatividade das mesmas. Então, a partir da abordagem proposta, espera-se que as imagens geradas sejam as mais informativas para o processo de aprendizado do classificador.

6.3 EXPERIMENTOS E RESULTADOS

Para avaliação da abordagem proposta foram definidos 4 fluxos de operação. A Figura 39 ilustra os 4 fluxos propostos.

- **GAN_Flow**: fluxo básico que não envolve o uso de técnicas de aprendizado ativo, ou seja, não inclui as etapas opcionais 4 e 6 da Figura 39.
- **AL-GAN_Flow**: fluxo que envolve o uso de técnicas de aprendizado ativo apenas na etapa que antecede o treinamento das GANs, ou seja, inclui a etapa opcional 4 da Figura 36.
- **GAN-AL_Flow**: fluxo que envolve o uso de técnicas de aprendizado ativo apenas na etapa seguinte ao treinamento das GANs, ou seja, inclui a etapa opcional 6 da Figura 36.
- **AL-GAN-AL_Flow**: fluxo que envolve o uso de técnicas de aprendizado ativo nas etapas pré e pós o treinamento das GANs, ou seja, inclui as etapas opcionais 4 e 6 da Figura 36.

Na Etapa 1 do *framework* proposto, foi realizada a separação de 20% dos dados de imagens reais para testes e os 80% remanescentes encontram-se disponíveis para serem utilizados no treinamento das GANs para geração das imagens sintéticas.

Diferentes estratégias de aprendizado ativo e classificadores podem ser considerados, nas Etapas 4 e/ou 6 do *framework* proposto, para seleção de imagens mais informativas para geração das imagens sintéticas e/ou para seleção de imagens sintéticas mais informativas. Além disso, diferentes quantidades de amostras podem ser selecionadas pelas estratégias de aprendizado ativo. A quantidade de amostras selecionadas foi definida como sendo 30% de amostras mais informativas, pois com tal quantidade é possível obter resultados equivalentes (por meio de estratégias de aprendizado ativo) quando comparado com o treinamento considerando o conjunto completo (vide Capítulo 5).

A Tabela 12 apresenta as acurácias obtidas pela estratégia MS-CLU, utilizando o classificador SVM para cada fluxo de operação do *framework* proposto (**GAN_Flow**, **AL-GAN_Flow**, **GAN-AL_Flow** e **AL-GAN-AL_Flow**) e para cada conjunto de enteroparasitos. Foram consideradas 1.000 imagens sintéticas geradas pela arquitetura BEGAN para cada classe

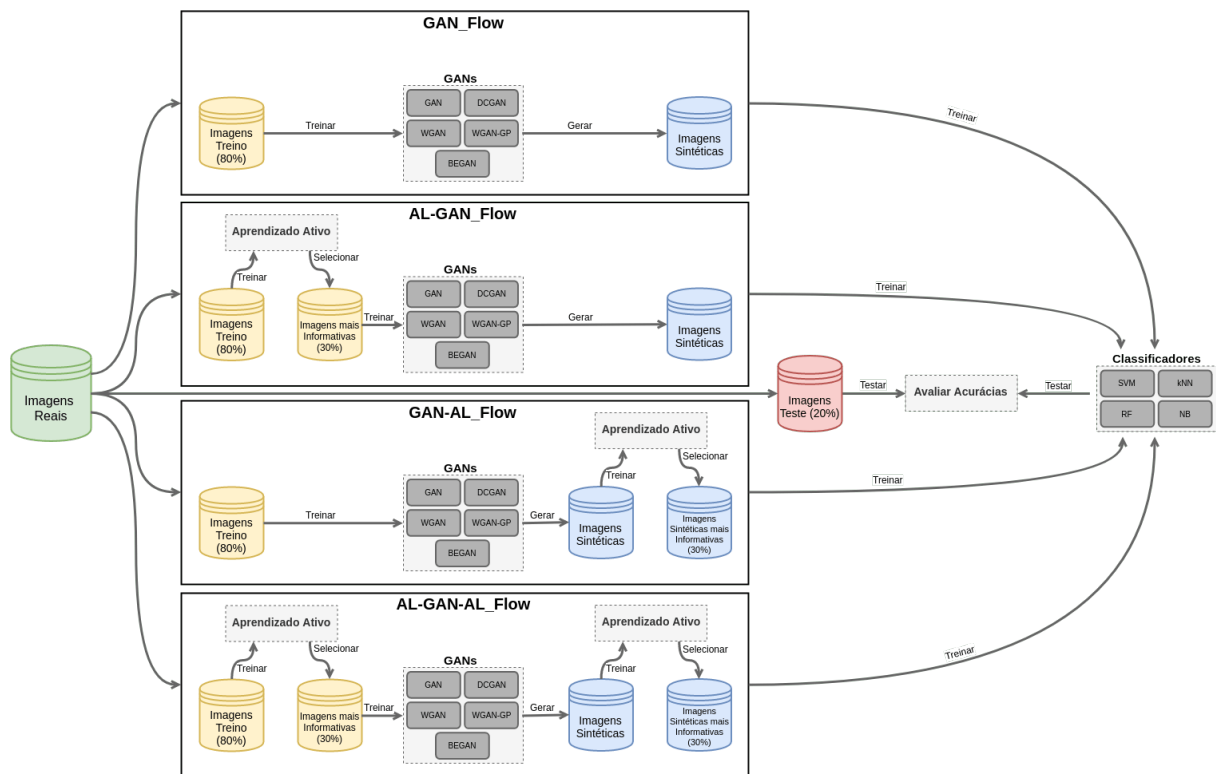


Figura 39: Fluxos do *framework* proposto.

presente nos conjuntos. Por exemplo, dado que o conjunto EGG-8 contém 8 classes, foram geradas 8.000 imagens sintéticas (1.000 para cada classe).

Tabela 12: Acurácias obtidas pela estratégia MS-CLU, utilizando o classificador SVM para cada fluxo de operação do *framework* proposto (GAN_Flow, AL-GAN_Flow, GAN-AL_Flow e AL-GAN-AL_Flow) e para cada conjunto de enteroparasitos.

	GAN_Flow	AL-GAN_Flow	GAN-AL_Flow	AL-GAN-AL_Flow
LAR-2	17.43	14.11	16.22	13.17
EGG-8	18.54	16.53	17.31	12.55
EGG-9	4.56	4.76	5.74	4.77
PRO-6	24.39	17.39	23.95	19.98
PRO-7	5.34	3.11	5.58	3.44
PAR-15	8.98	8.85	9.70	6.91
PAR-16	6.12	5.13	6.22	4.87

Comparando os fluxos que envolvem a inclusão de estratégias de aprendizado ativo para a *geração das imagens sintéticas* (AL-GAN_FLOW e AL-GAN-AL_FLOW) com os fluxos que não utilizam aprendizado ativo antes do treinamento das GANs para geração de imagens sintéticas (GAN_FLOW e GAN-AL_FLOW), pode-se observar que os resultados com os fluxos GAN_FLOW e GAN-AL_FLOW apresentam valores de acurácias **discretamente** superiores aos dos fluxos AL-GAN_FLOW e AL-GAN-AL_FLOW. Por exemplo, considerando o conjunto PAR-15, GAN-Flow e AL-GAN_Flow apresentam valores de acurácias similares

(8.98% e 8.85%, respectivamente).

É importante ressaltar que os fluxos que consideram estratégias de aprendizado ativo (**AL-GAN_Flow**, **GAN-AL_Flow** e **AL-GAN-AL_Flow**) não consideram o conjunto completo (seja de imagens reais e/ou sintéticas). Tais fluxos consideram apenas 30% das imagens (reais e/ou sintéticas). Nesse sentido, pode-se observar o desempenho superior da abordagem proposta, a qual consegue obter acurácias similares com um pequeno conjunto rotulado contendo imagens mais informativas, por meio de estratégias de aprendizado ativo.

A partir da redução na quantidade de imagens rotuladas (i.e. redução de custo e de necessidade de anotação realizado por um especialista), diminui-se também drasticamente o tempo computacional para ambos processos de síntese e de classificação das imagens. Por exemplo, considerando que o tempo médio da etapa de treinamento na arquitetura BEGAN seja de 25 dias para o conjunto completo PAR-15, utilizando os fluxos que consideram estratégias de aprendizado ativo (**AL-GAN_FLOW** e **AL-GAN-AL_FLOW**) o tempo é reduzido para aproximadamente 8 dias. Ao selecionar apenas os 30% de amostras mais informativas reduz-se quase que na mesma proporção o tempo de treinamento das GANs.

Além do processo de síntese de imagens, outra contribuição apresentada pela abordagem proposta refere-se à inclusão de estratégias de aprendizado ativo para o processo de *classificação das imagens* (fluxos **GAN-AL_Flow** e **AL-GAN-AL_Flow**), selecionando as imagens sintéticas mais informativas para o aprendizado do classificador. Nesse caso, a abordagem proposta mostra-se também mais eficaz e eficiente. É possível observar ganhos em relação à acurácia e ao tempo de treinamento do classificador final. Por exemplo, considerando o conjunto PAR-15, para os fluxos **GAN_Flow** e **AL-GAN_FLOW** (i.e. sem selecionar as amostras sintéticas mais informativas por meio de estratégias de aprendizado ativo) obtém-se valores de acurácias 8.29% e 5.85%, respectivamente, utilizando 15.000 imagens sintéticas (1.000 para cada classe). Já para os fluxos **GAN-AL_Flow** e **AL-GAN-AL_FLOW** obtém-se valores de acurácias superiores (8.89% e 6.83%, respectivamente), bem como custos menores de treinamento do classificador final, devido à redução no conjunto de treinamento (4.500 amostras).

Vale destacar que os valores (baixos) de acurácia estão relacionados à qualidade das imagens sintéticas geradas pelas GANs. Apesar de não ser o foco do presente projeto, foram realizados experimentos de forma a avaliar diferentes arquiteturas (GAN, DCGAN, WGAN, WGAN-GP e BEGAN) para geração das imagens sintéticas. Melhorias nos valores finais de acurácias podem ser obtidas por meio de arquiteturas de GANs mais robustas (XU et al., 2018; SHMELKOV et al., 2018). No entanto, tais arquiteturas demandam um grande poder computacional, o qual não estava disponível durante a realização desse trabalho. Outras alternativas

para síntese de imagens podem ser exploradas como por exemplo *autoencoders* (BOURLARD; KAMP, 1988) e suas variações (KINGMA; WELLING, 2013; HIGGINS et al., 2016; HUANG et al., 2018). Outro ponto de melhoria seria a utilização de máscaras de segmentação para remoção da informação de fundo das imagens dos conjuntos de enteroparasitos, pois a presença dessas informações de fundo pode dificultar o treinamento das GANs.

Embora os valores finais de acurácias não sejam elevados, a partir dos experimentos e resultados apresentados, é possível observar as vantagens apresentadas pela abordagem proposta, analisando os ganhos (acurácias e tempos computacionais) apresentados pela inclusão de estratégias de aprendizado ativo, as quais conseguem de fato selecionar amostras mais informativas tanto para a *síntese* como para a *classificação* das imagens.

7 CONCLUSÕES

É fundamental o desenvolvimento de abordagens de aprendizado robustas, considerando os problemas relacionados à i-) escassez de imagens, dada a dificuldade de aquisição de imagens reais; ii-) pouca quantidade de imagens rotuladas, dada a dificuldade de anotação por especialistas; iii-) necessidade de um grande conjunto de imagens rotuladas, dada a dificuldade de treinamento e obtenção de um classificador de padrões robusto. Portanto, este trabalho propõe uma nova abordagem para anotação de imagens, com enfoque no domínio de enteroparasitos, explorando desde técnicas para obtenção dos melhores descritores para as imagens até estratégias de aprendizado ativo para síntese e classificação das imagens.

Uma vez que a extração de características das imagens é determinante para a qualidade final dos algoritmos de aprendizado, inicialmente, foram elaborados experimentos (Capítulo 4) (ALVES et al., 2017; BRESSAN et al., 2018; VALERIO et al., 2019) comparando características extraídas por meio de extratores tradicionais e por meio de técnicas de aprendizado profundo (redes neurais convolucionais e *Transfer Learning*).

Após a obtenção das características que melhor descrevem cada conjunto de imagens, novas estratégias de aprendizado ativo foram propostas e avaliadas (Capítulo 5) (ALVES et al., 2019a, 2019b), de forma a aplicar as estratégias que selecionam as amostras mais informativas para os processos de síntese e classificação das imagens. As estratégias propostas apresentaram resultados superiores em alguns cenários e métricas considerados.

Para avaliar a abordagem de aprendizado proposta (Capítulo 6) (ALVES et al., 2019c) foi analisado o impacto da inclusão de estratégias de aprendizado ativo, nos processos mútuos de geração e seleção de imagens sintéticas mais informativas, a fim de obter conjuntamente sintetizadores de imagens e classificadores de padrões mais efetivos e eficientes.

Embora não seja o foco do presente projeto, foram realizados experimentos prévios envolvendo o uso de diferentes arquiteturas de GANs para geração das imagens sintéticas, bem como de métricas de avaliação das GANs. Esforços foram despendidos, tanto na escolha de arquitetura, implementações, treinamento e ajustes de hiper-parâmetros das GANs. No entanto,

nenhum conjunto de amostras sintéticas do domínio de enteroparasitos conseguiu agregar a qualidade visual necessária com o bom desempenho segundo as métricas consideradas. Por meio dos experimentos, foi demonstrado que os resultados obtidos são consistentes aos reportados na literatura. Existe um grau de complexidade maior no domínio de enteroparasitos quando comparado a outros conjuntos mais simples (como por exemplo o CIFAR-10), em que os resultados foram mais promissores.

A qualidade das imagens sintéticas geradas pelas GANs impacta significativamente no desempenho (medidas de acurácias) do classificador. Melhorias na qualidade das imagens sintéticas geradas (e, conseqüentemente em acurácias elevadas apresentadas pelo classificador) podem ser obtidas por meio de arquiteturas de GANs mais robustas. Em contrapartida, são necessários recursos computacionais adequados. Outras alternativas podem ser exploradas como por exemplo a utilização de *autoencoders* e/ou suas variações. Outra possibilidade de melhoria seria investigar a utilização das imagens dos conjuntos de enteroparasitos com a remoção das informações de fundo das imagens, dado que a presença de tais informações pode ter impactado negativamente no treinamento das GANs.

Apesar dos resultados (valores finais de acurácias) obtidos não tenham sido elevados, é possível observar as vantagens apresentadas pela abordagem proposta. A partir dos experimentos e resultados apresentados, nota-se que as estratégias de aprendizado ativo conseguem selecionar amostras mais informativas para ambos os processos de síntese e de classificação das imagens. A abordagem proposta mostra-se eficaz e eficiente, atingindo acurácias similares e tempos computacionais menores (ao reduzir significativamente a quantidade de imagens rotuladas e a necessidade de esforço do especialista no processo de anotação).

Portanto, a abordagem proposta, com a inclusão de estratégias de aprendizado ativo, mostra-se promissora, com valiosas contribuições para a **síntese**: resolvendo problemas de desbalanceamento de dados oriundos de aplicações reais, ao possibilitar a geração de imagens sintéticas mais informativas de classes que apresentam pouca quantidade de amostras; e para a **classificação**: acelerando e melhorando os processos de aprendizado e anotação de grandes bases de imagens, ao auxiliar especialistas indicando imagens mais informativas para construção do classificador.

REFERÊNCIAS

- ABEDINI, M. et al. A generalized framework for medical image classification and recognition. **IBM Journal of Research and Development**, v. 59, n. 2/3, 2015. Disponível em: <<https://doi.org/10.1147/JRD.2015.2390017>>.
- ALAJLAN, N. et al. Large-scale image classification using active learning. **IEEE Geosci. Remote Sensing Lett.**, v. 11, n. 1, p. 259–263, 2014. Disponível em: <<https://doi.org/10.1109/LGRS.2013.2255258>>.
- ALIPANAHI, B. et al. Predicting the sequence specificities of dna-and rna-binding proteins by deep learning. **Nature biotechnology**, Nature Publishing Group, v. 33, n. 8, p. 831, 2015.
- ALVES, D. H. et al. Biolearner: A tool for interactive analysis and visualization of active learning strategies for bioimages (to be submitted). **Briefings in bioinformatics**, Oxford University Press, 2019.
- ALVES, D. H. et al. Cluster-based active learning strategies for bioimages classification: A case study on human intestinal parasites (to be submitted). **Medical Image Analysis**, Elsevier, 2019.
- ALVES, D. H. et al. Does gans learn better with most informative samples? (to be submitted). **IEEE Transactions on Neural Networks and Learning Systems**, IEEE, 2019.
- ALVES, D. H. A. et al. Going deeper on bioimages classification: A plant leaf dataset case study. In: **Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications - 22nd Iberoamerican Congress, CIARP 2017, Valparaíso, Chile, November 7-10, 2017, Proceedings**. [s.n.], 2017. p. 36–44. Disponível em: <https://doi.org/10.1007/978-3-319-75193-1_5>.
- AMODEI, D. et al. Deep speech 2 : End-to-end speech recognition in english and mandarin. In: **Proceedings of the 33rd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016**. [s.n.], 2016. p. 173–182. Disponível em: <<http://jmlr.org/proceedings/papers/v48/amodei16.html>>.
- ASGHAR, N. et al. Deep active learning for dialogue generation. In: **Proceedings of the 6th Joint Conference on Lexical and Computational Semantics, *SEM @ACM 2017, Vancouver, Canada, August 3-4, 2017**. [s.n.], 2017. p. 78–83. Disponível em: <<https://doi.org/10.18653/v1/S17-1008>>.
- BAI, Y. et al. A novel feature extraction method using pyramid histogram of orientation gradients for smile recognition. In: **Proceedings of the International Conference on Image Processing, ICIP 2009, 7-10 November 2009, Cairo, Egypt**. [s.n.], 2009. p. 3305–3308. Disponível em: <<https://doi.org/10.1109/ICIP.2009.5413938>>.
- BARRATT, S.; SHARMA, R. A note on the inception score. **arXiv preprint arXiv:1801.01973**, 2018.

BENGIO, Y.; SIMARD, P. Y.; FRASCONI, P. Learning long-term dependencies with gradient descent is difficult. **IEEE Trans. Neural Networks**, v. 5, n. 2, p. 157–166, 1994. Disponível em: <<https://doi.org/10.1109/72.279181>>.

BERTHELOT, D.; SCHUMM, T.; METZ, L. BEGAN: boundary equilibrium generative adversarial networks. **CoRR**, abs/1703.10717, 2017. Disponível em: <<http://arxiv.org/abs/1703.10717>>.

BOJARSKI, M. et al. End to end learning for self-driving cars. **CoRR**, abs/1604.07316, 2016. Disponível em: <<http://arxiv.org/abs/1604.07316>>.

BOURLARD, H.; KAMP, Y. Auto-association by multilayer perceptrons and singular value decomposition. **Biological cybernetics**, Springer, v. 59, n. 4-5, p. 291–294, 1988.

BREIMAN, L. Random forests. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>.

BRESSAN, R. S. et al. Doctor: the role of deep features in content-based mammographic image retrieval. In: IEEE. **2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2018. p. 158–163.

BRESSAN, R. S.; BUGATTI, P. H.; SAITO, P. T. Breast cancer diagnosis through active learning in content-based image retrieval. **Neurocomputing**, Elsevier, v. 357, p. 1–10, 2019.

BRESSAN, R. S. et al. Exploring active learning based on representativeness and uncertainty for biomedical data classification. **IEEE journal of biomedical and health informatics**, IEEE, 2018.

BROCK, A.; DONAHUE, J.; SIMONYAN, K. Large scale gan training for high fidelity natural image synthesis. **arXiv preprint arXiv:1809.11096**, 2018.

BRYANT, F. B.; YARNOLD, P. R. Principal-components analysis and exploratory and confirmatory factor analysis. American Psychological Association, 1995.

BURLINA, P. et al. Comparing humans and deep learning performance for grading AMD: A study in using universal deep features and transfer learning for automated AMD analysis. **Comp. in Bio. and Med.**, v. 82, p. 80–86, 2017. Disponível em: <<https://doi.org/10.1016/j.combiomed.2017.01.018>>.

CHANDRARATNE, M.; KULASIRI, D.; SAMARASINGHE, S. Classification of lamb carcass using machine vision: Comparison of statistical and neural network analyses. **Journal of Food Engineering**, v. 82, n. 1, p. 26 – 34, 2007. ISSN 0260-8774. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0260877407000404>>.

CHO, K. et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. p. 1724–1734, 2014. Disponível em: <<http://aclweb.org/anthology/D/D14/D14-1179.pdf>>.

CHOWDHURY, A.; BIANCO, S.; BISWAS, S. Active deep learning reduces annotation burden in automatic cell segmentation. **bioRxiv**, Cold Spring Harbor Laboratory, 2017. Disponível em: <<https://www.biorxiv.org/content/early/2017/10/29/211060>>.

COVER, T. M.; HART, P. E. Nearest neighbor pattern classification. **IEEE Trans. Information Theory**, v. 13, n. 1, p. 21–27, 1967. Disponível em: <<https://doi.org/10.1109/TIT.1967.1053964>>.

DASKALAKIS, C.; GOLDBERG, P. W.; PAPADIMITRIOU, C. H. The complexity of computing a nash equilibrium. **Commun. ACM**, v. 52, n. 2, p. 89–97, 2009. Disponível em: <<http://doi.acm.org/10.1145/1461928.1461951>>.

DENTON, E. L. et al. Deep generative image models using a laplacian pyramid of adversarial networks. In: **Advances in Neural Information Processing Systems 28: Annual Conference on Neural Information Processing Systems 2015, December 7-12, 2015, Montreal, Quebec, Canada**. [s.n.], 2015. p. 1486–1494. Disponível em: <<http://papers.nips.cc/paper/5773-deep-generative-image-models-using-a-laplacian-pyramid-of-adversarial-networks>>.

DOMINGOS, P. M.; PAZZANI, M. J. On the optimality of the simple bayesian classifier under zero-one loss. **Machine Learning**, v. 29, n. 2-3, p. 103–130, 1997. Disponível em: <<https://doi.org/10.1023/A:1007413511361>>.

DU, X.; ZHU, F. A novel principal components analysis (pca) method for energy absorbing structural design enhanced by data mining. **Advances in Engineering Software**, Elsevier, v. 127, p. 17–27, 2019.

GAL, Y.; ISLAM, R.; GHAHRAMANI, Z. Deep bayesian active learning with image data. p. 1183–1192, 2017. Disponível em: <<http://proceedings.mlr.press/v70/gal17a.html>>.

GOËAU, H. et al. The ImageCLEF 2012 Plant Identification Task. In: **CLEF'2012: Conference and Labs of the Evaluation Forum**. Rome, Italy: [s.n.], 2012. Disponível em: <<https://hal.inria.fr/hal-00960918>>.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. **Deep learning**. [S.l.]: MIT press, 2016.

GOODFELLOW, I. J. et al. Generative adversarial nets. In: **Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada**. [s.n.], 2014. p. 2672–2680. Disponível em: <<http://papers.nips.cc/paper/5423-generative-adversarial-nets>>.

GULRAJANI, I. et al. Improved training of wasserstein gans. p. 5769–5779, 2017. Disponível em: <<http://papers.nips.cc/paper/7159-improved-training-of-wasserstein-gans>>.

HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. **Journal of the Royal Statistical Society. Series C (Applied Statistics)**, JSTOR, v. 28, n. 1, p. 100–108, 1979.

HE, K. et al. Deep residual learning for image recognition. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016**. [s.n.], 2016. p. 770–778. Disponível em: <<https://doi.org/10.1109/CVPR.2016.90>>.

HE, K. et al. Identity mappings in deep residual networks. In: **Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part IV**. [s.n.], 2016. p. 630–645. Disponível em: <https://doi.org/10.1007/978-3-319-46493-0_38>.

- HEARST, M. A. et al. Support vector machines. **IEEE Intelligent Systems and their Applications**, IEEE, v. 13, n. 4, p. 18–28, 1998.
- HEUSEL, M. et al. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2017. p. 6626–6637.
- HIGGINS, I. et al. Early visual concept learning with unsupervised deep learning. **arXiv preprint arXiv:1606.05579**, 2016.
- HINTON, G. E.; VINYALS, O.; DEAN, J. Distilling the knowledge in a neural network. **CoRR**, abs/1503.02531, 2015. Disponível em: <<http://arxiv.org/abs/1503.02531>>.
- HU, J.; SHEN, L.; SUN, G. Squeeze-and-excitation networks. **CoRR**, abs/1709.01507, 2017. Disponível em: <<http://arxiv.org/abs/1709.01507>>.
- HUANG, H. et al. Introvae: Introspective variational autoencoders for photographic image synthesis. In: **Advances in Neural Information Processing Systems**. [S.l.: s.n.], 2018. p. 52–63.
- HUANG, J. et al. Spatial color indexing and applications. **International Journal of Computer Vision**, v. 35, n. 3, p. 245–268, 1999. Disponível em: <<https://doi.org/10.1023/A:1008108327226>>.
- HUANG, S.-J.; JIN, R.; ZHOU, Z.-H. Active learning by querying informative and representative examples. In: **Advances in neural information processing systems**. [S.l.: s.n.], 2010. p. 892–900.
- HUSSEIN, A.; GABER, M. M.; ELYAN, E. Deep active learning for autonomous navigation. In: **Engineering Applications of Neural Networks - 17th International Conference, EANN 2016, Aberdeen, UK, September 2-5, 2016, Proceedings**. [s.n.], 2016. p. 3–17. Disponível em: <https://doi.org/10.1007/978-3-319-44188-7_1>.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: **Proceedings of the 32nd International Conference on Machine Learning, ICML 2015, Lille, France, 6-11 July 2015**. [s.n.], 2015. p. 448–456. Disponível em: <<http://jmlr.org/proceedings/papers/v37/ioffe15.html>>.
- KAREGOWDA, A. G.; MANJUNATH, A.; JAYARAM, M. Comparative study of attribute selection using gain ratio and correlation based feature selection. **International Journal of Information Technology and Knowledge Management**, v. 2, n. 2, p. 271–277, 2010.
- KARRAS, T. et al. Progressive growing of gans for improved quality, stability, and variation. **arXiv preprint arXiv:1710.10196**, 2017.
- KASUTANI, E.; YAMADA, A. The mpeg-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. In: IEEE. **ICIP**. [S.l.], 2001. p. 674–677.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. **CoRR**, abs/1412.6980, 2014. Disponível em: <<http://arxiv.org/abs/1412.6980>>.

- KINGMA, D. P.; WELLING, M. Auto-encoding variational bayes. **arXiv preprint arXiv:1312.6114**, 2013.
- KOHONEN, T. The self-organizing map. **Neurocomputing**, v. 21, n. 1-3, p. 1–6, 1998. Disponível em: <[https://doi.org/10.1016/S0925-2312\(98\)00030-7](https://doi.org/10.1016/S0925-2312(98)00030-7)>.
- KOUROU, K. et al. Machine learning applications in cancer prognosis and prediction. **Computational and Structural Biotechnology Journal**, v. 13, p. 8 – 17, 2015. ISSN 2001-0370. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2001037014000464>>.
- KRIZHEVSKY, A.; HINTON, G. **Learning multiple layers of features from tiny images**. [S.l.], 2009.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: **Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States**. [s.n.], 2012. p. 1106–1114. Disponível em: <<http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks>>.
- LECUN, Y.; BENGIO, Y.; HINTON, G. E. Deep learning. **Nature**, v. 521, n. 7553, p. 436–444, 2015. Disponível em: <<https://doi.org/10.1038/nature14539>>.
- LECUN, Y. et al. Gradient-based learning applied to document recognition. **Proceedings of the IEEE**, IEEE, v. 86, n. 11, p. 2278–2324, 1998.
- LIU, B. **Sentiment Analysis and Opinion Mining**. Morgan & Claypool Publishers, 2012. (Synthesis Lectures on Human Language Technologies). Disponível em: <<https://doi.org/10.2200/S00416ED1V01Y201204HLT016>>.
- LOPEZ-PAZ, D.; OQUAB, M. Revisiting classifier two-sample tests. **arXiv preprint arXiv:1610.06545**, 2016.
- NIXON, M. S.; AGUADO, A. S. **Feature extraction & image processing for computer vision**. [S.l.]: Academic Press, 2012.
- PAN, S. J.; YANG, Q. et al. A survey on transfer learning. **IEEE Transactions on knowledge and data engineering**, Institute of Electrical and Electronics Engineers, Inc., 345 E. 47 th St. NY ..., v. 22, n. 10, p. 1345–1359, 2010.
- PAPA, J. P.; FALCÃO, A. X.; SUZUKI, C. T. N. Supervised pattern classification based on optimum-path forest. **Int. J. Imaging Systems and Technology**, v. 19, n. 2, p. 120–131, 2009. Disponível em: <<https://doi.org/10.1002/ima.20188>>.
- PRICE, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature genetics**, Nature Publishing Group, v. 38, n. 8, p. 904, 2006.
- QUINLAN, J. R. **C4. 5: programs for machine learning**. [S.l.]: Elsevier, 2014.
- RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. **CoRR**, abs/1511.06434, 2015. Disponível em: <<http://arxiv.org/abs/1511.06434>>.

RAWAT, W.; WANG, Z. Deep convolutional neural networks for image classification: A comprehensive review. **Neural Computation**, v. 29, n. 9, p. 2352–2449, 2017. Disponível em: <https://doi.org/10.1162/neco_a_00990>.

REED, S. E. et al. Generative adversarial text to image synthesis. p. 1060–1069, 2016. Disponível em: <<http://jmlr.org/proceedings/papers/v48/reed16.html>>.

REKER, D.; SCHNEIDER, G. Active-learning strategies in computer-assisted drug discovery. **Drug discovery today**, Elsevier, v. 20, n. 4, p. 458–465, 2015.

ROCHA, L. M.; CAPPABIANCO, F. A. M.; FALCÃO, A. X. Data clustering as an optimum-path forest problem with applications in image analysis. **Int. J. Imaging Systems and Technology**, v. 19, n. 2, p. 50–68, 2009. Disponível em: <<https://doi.org/10.1002/ima.20191>>.

ROSENBLATT, F. The perceptron: A probabilistic model for information storage and organization in the brain. **Psychological review**, American Psychological Association, v. 65, n. 6, p. 386, 1958.

RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **nature**, Nature Publishing Group, v. 323, n. 6088, p. 533, 1986.

RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. **International Journal of Computer Vision**, v. 115, n. 3, p. 211–252, 2015. Disponível em: <<https://doi.org/10.1007/s11263-015-0816-y>>.

SABOUR, S.; FROSST, N.; HINTON, G. E. Dynamic routing between capsules. In: **Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA**. [s.n.], 2017. p. 3859–3869. Disponível em: <<http://papers.nips.cc/paper/6975-dynamic-routing-between-capsules>>.

SAITO, P. T. M. Drop: A data reduction and organization paradigm and its application in image analysis. **ELCVIA: electronic letters on computer vision and image analysis**, v. 14, n. 3, 2015.

SAITO, P. T. M. et al. Active semi-supervised learning using optimum-path forest. In: **22nd International Conference on Pattern Recognition, ICPR 2014, Stockholm, Sweden, August 24-28, 2014**. [s.n.], 2014b. p. 3798–3803. Disponível em: <<https://doi.org/10.1109/ICPR.2014.652>>.

SAITO, P. T. M. et al. Improving active learning with sharp data reduction. **WSCG'2012**, Cite-seer, 2012.

SAITO, P. T. M. et al. A data reduction and organization approach for efficient image annotation. In: **Proceedings of the 28th Annual ACM Symposium on Applied Computing, SAC '13, Coimbra, Portugal, March 18-22, 2013**. [s.n.], 2013. p. 53–57. Disponível em: <<http://doi.acm.org/10.1145/2480362.2480374>>.

SAITO, P. T. M. et al. An active learning paradigm based on a priori data reduction and organization. **Expert Syst. Appl.**, v. 41, n. 14, p. 6086–6097, 2014a. Disponível em: <<https://doi.org/10.1016/j.eswa.2014.04.007>>.

SAITO, P. T. M. et al. Robust active learning for the diagnosis of parasites. **Pattern Recognition**, v. 48, n. 11, p. 3572–3583, 2015. Disponível em: <<https://doi.org/10.1016/j.patcog.2015.05.020>>.

SALIMANS, T. et al. Improved techniques for training gans. In: **Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain**. [s.n.], 2016. p. 2226–2234. Disponível em: <<http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans>>.

SETTLES, B. Active learning. **Synthesis Lectures on Artificial Intelligence and Machine Learning**, Morgan & Claypool Publishers, v. 6, n. 1, p. 1–114, 2012.

SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. [S.l.]: Cambridge university press, 2014.

SHEN, Y. et al. Deep active learning for named entity recognition. Association for Computational Linguistics, p. 252–256, 2017. Disponível em: <<https://aclanthology.info/papers/W17-2630/w17-2630>>.

SHIN, H. et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. **IEEE Trans. Med. Imaging**, v. 35, n. 5, p. 1285–1298, 2016. Disponível em: <<https://doi.org/10.1109/TMI.2016.2528162>>.

SHMELKOV, K.; SCHMID, C.; ALAHARI, K. How good is my gan? In: **Proceedings of the European Conference on Computer Vision (ECCV)**. [S.l.: s.n.], 2018. p. 213–229.

SRIVASTAVA, N. et al. Dropout: a simple way to prevent neural networks from overfitting. **Journal of Machine Learning Research**, v. 15, n. 1, p. 1929–1958, 2014. Disponível em: <<http://dl.acm.org/citation.cfm?id=2670313>>.

SUZUKI, C. T. N. et al. Automated diagnosis of human intestinal parasites using optical microscopy images. In: **10th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, ISBI 2013, 7-11 April, 2013, San Francisco, CA, USA, Proceedings**. [s.n.], 2013. p. 460–463. Disponível em: <<https://doi.org/10.1109/ISBI.2013.6556511>>.

SZEGEDY, C. et al. Inception-v4, inception-resnet and the impact of residual connections on learning. In: SINGH, S. P.; MARKOVITCH, S. (Ed.). **Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA**. AAAI Press, 2017. p. 4278–4284. Disponível em: <<http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14806>>.

SZEGEDY, C. et al. Going deeper with convolutions. In: **IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015**. [s.n.], 2015. p. 1–9. Disponível em: <<https://doi.org/10.1109/CVPR.2015.7298594>>.

SZEGEDY, C. et al. Rethinking the inception architecture for computer vision. In: **2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016**. [s.n.], 2016. p. 2818–2826. Disponível em: <<https://doi.org/10.1109/CVPR.2016.308>>.

TAYLOR, M. E.; STONE, P. Transfer learning for reinforcement learning domains: A survey. **Journal of Machine Learning Research**, v. 10, n. Jul, p. 1633–1685, 2009.

UNTERTHINER, T. et al. Coulomb gans: Provably optimal nash equilibria via potential fields. **CoRR**, abs/1708.08819, 2017. Disponível em: <<http://arxiv.org/abs/1708.08819>>.

VALERIO, L. M. et al. Deepmammo: Deep transfer learning for lesion classification of mammographic images. In: IEEE. **2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)**. [S.l.], 2019. p. 447–452.

WANG, K. et al. Cost-effective active learning for deep image classification. **IEEE Trans. Circuits Syst. Video Techn.**, v. 27, n. 12, p. 2591–2600, 2017. Disponível em: <<https://doi.org/10.1109/TCSVT.2016.2589879>>.

WON, C. S.; PARK, D. K.; PARK, S.-J. Efficient use of mpeg-7 edge histogram descriptor. **ETRI journal**, Electronics and Telecommunications Research Institute, v. 24, n. 1, p. 23–30, 2002.

XU, Q. et al. An empirical study on evaluation metrics of generative adversarial networks. **arXiv preprint arXiv:1806.07755**, 2018.

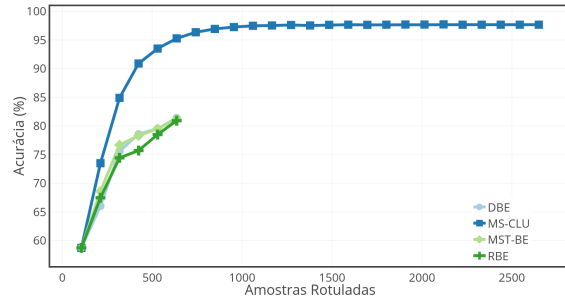
YANG, L. et al. Suggestive annotation: A deep active learning framework for biomedical image segmentation. In: **Medical Image Computing and Computer Assisted Intervention - MICCAI 2017 - 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III**. [s.n.], 2017. p. 399–407. Disponível em: <https://doi.org/10.1007/978-3-319-66179-7_46>.

YOSINSKI, J. et al. How transferable are features in deep neural networks? In: **Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada**. [s.n.], 2014. p. 3320–3328. Disponível em: <<http://papers.nips.cc/paper/5347-how-transferable-are-features-in-deep-neural-networks>>.

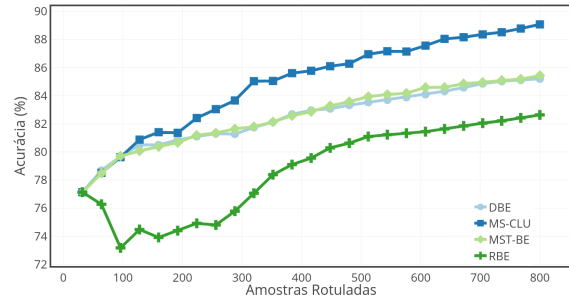
ZHU, J.; BENTO, J. Generative adversarial active learning. **CoRR**, abs/1702.07956, 2017. Disponível em: <<http://arxiv.org/abs/1702.07956>>.

ZHU, J. et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: **IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017**. [s.n.], 2017. p. 2242–2251. Disponível em: <<https://doi.org/10.1109/ICCV.2017.244>>.

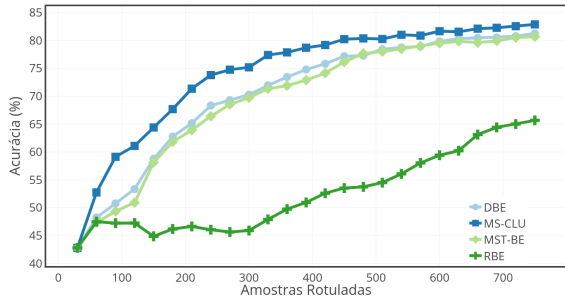
**APÊNDICE A – ACURÁCIAS ESTRATÉGIAS RBE, DBE, MST-BE E MS-CLU COM
CLASSIFICADOR SVM**



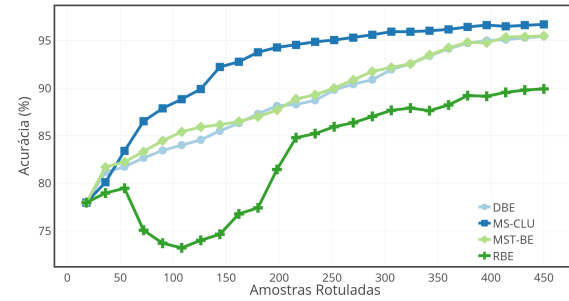
(a) LEA-53



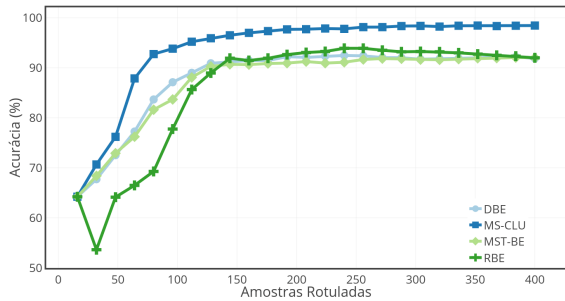
(b) PAR-16



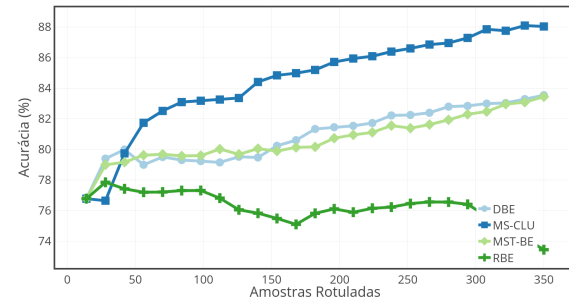
(c) PAR-15



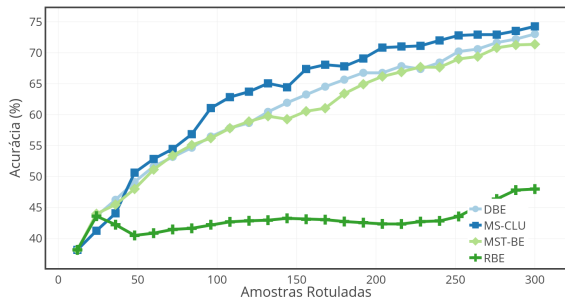
(d) EGG-9



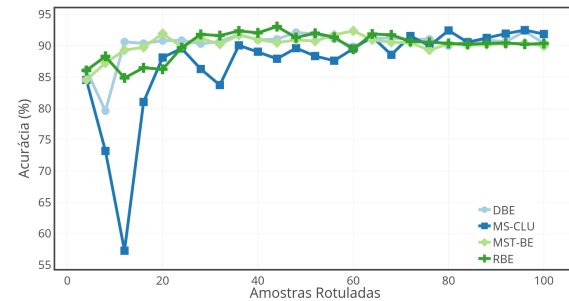
(e) EGG-8



(f) PRO-7



(g) PRO-6



(h) LAR-2

Figura 40: Médias das acurácias obtidas pelas estratégias de aprendizado ativo DBE, MS-CLU, MST-BE e RBE com o classificador SVM nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

**APÊNDICE B – DESEMPENHO DAS ESTRATÉGIAS MS-CLU, MST-CLU-DS,
MST-CLU-DDE, MST-CLU-RDS E RDS COM CLASSIFICADOR
K-NN**

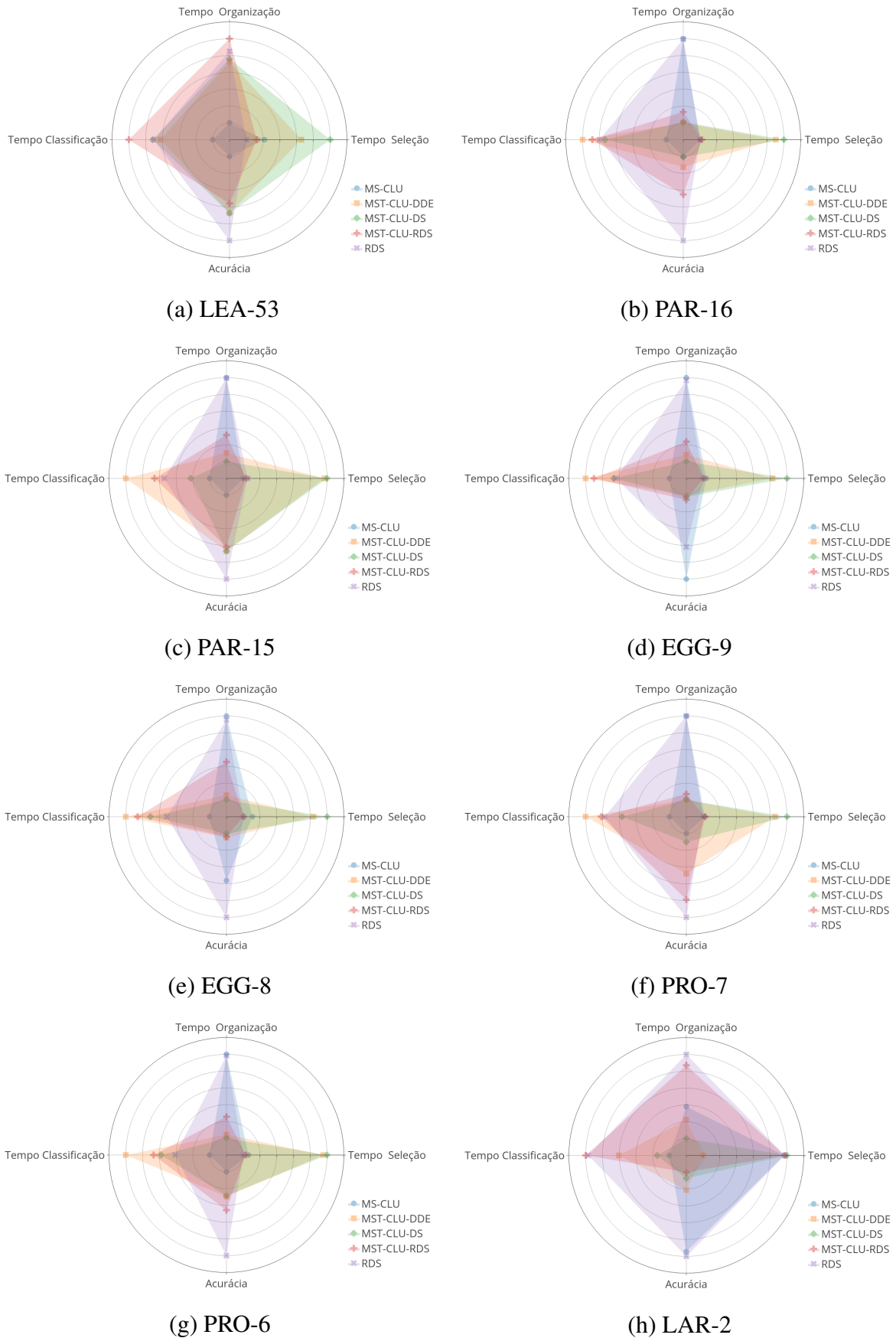


Figura 41: Médias de acurácias, tempo de organização, tempo de seleção e tempo de classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador k -NN nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

**APÊNDICE C – DESEMPENHO DAS ESTRATÉGIAS MS-CLU, MST-CLU-DS,
MST-CLU-DDE, MST-CLU-RDS E RDS COM CLASSIFICADOR RF**

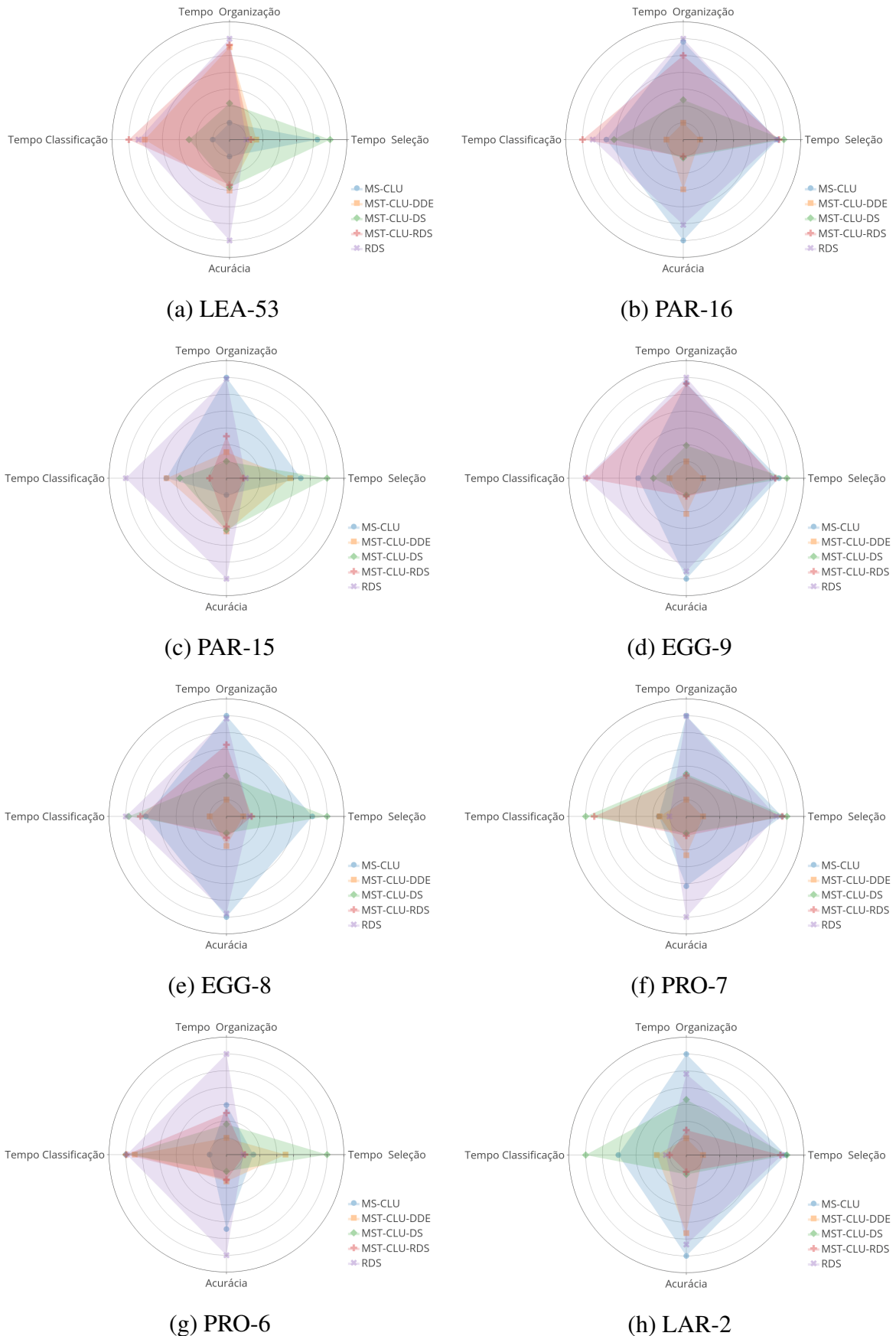


Figura 42: Médias de acurácias, tempo de organização, tempo de seleção e tempo de classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador RF nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

**APÊNDICE D – DESEMPENHO DAS ESTRATÉGIAS MS-CLU, MST-CLU-DS,
MST-CLU-DDE, MST-CLU-RDS E RDS COM CLASSIFICADOR NB**

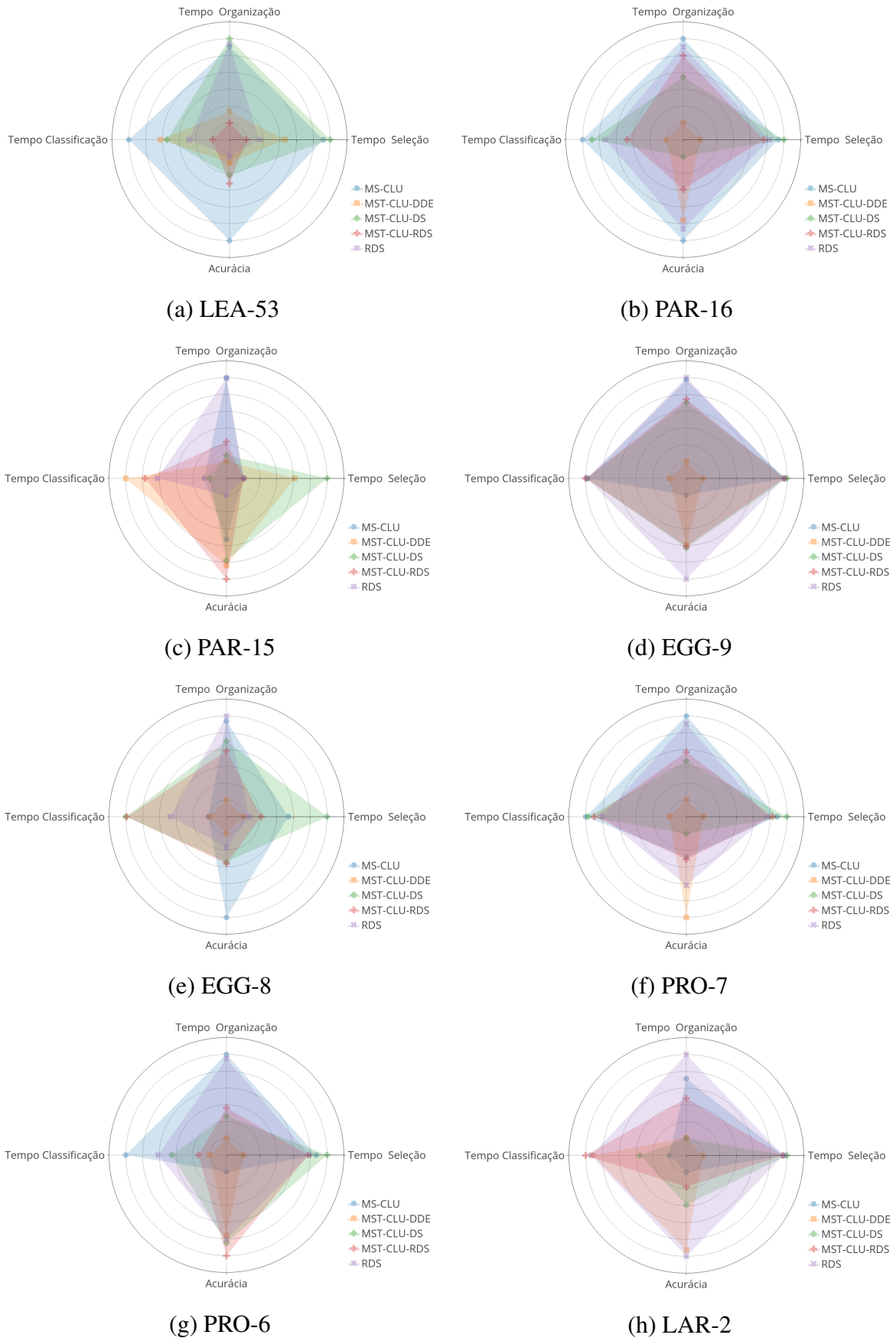
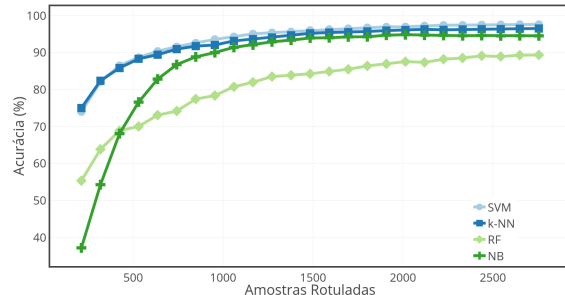
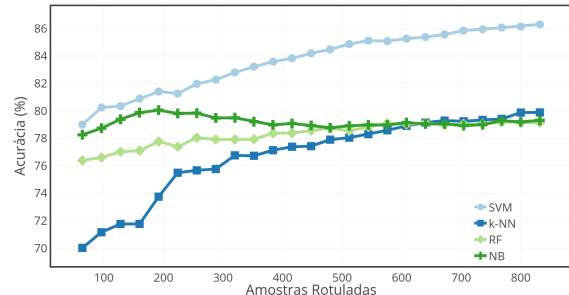


Figura 43: Médias de acurácias, tempo de organização, tempo de seleção e tempo de classificação obtidas pelas estratégias de aprendizado ativo MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

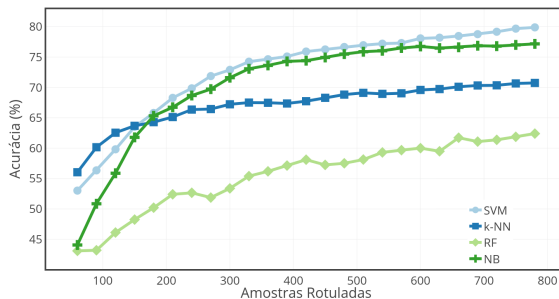
**APÊNDICE E – ACURÁCIAS DA ESTRATÉGIA MST-CLU-DDE COM
CLASSIFICADORES SVM, K-NN, RF E NB**



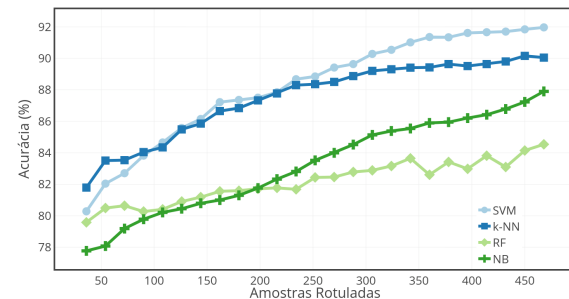
(a) LEA-53



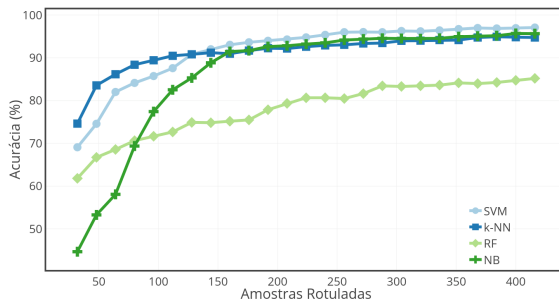
(b) PAR-16



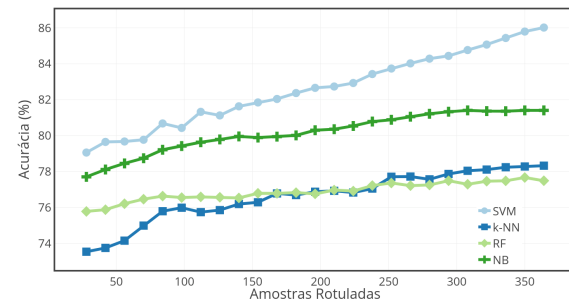
(c) PAR-15



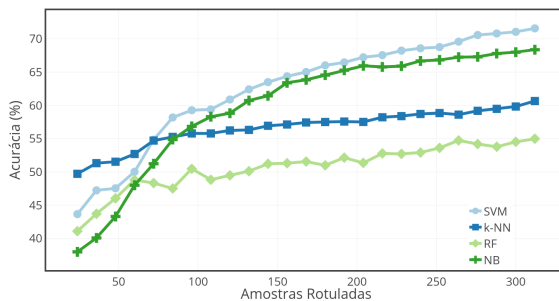
(d) EGG-9



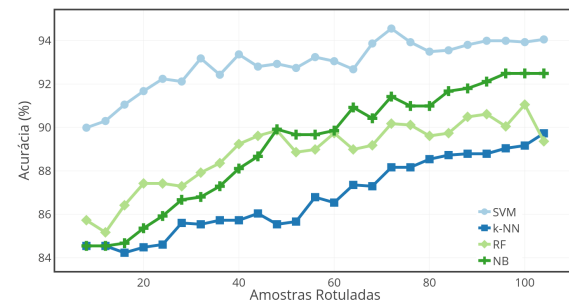
(e) EGG-8



(f) PRO-7



(g) PRO-6



(h) LAR-2

Figura 44: Médias das acurácias obtidas pela estratégia de aprendizado ativo MST-CLU-DDE com os classificadores SVM, *k*-NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

**APÊNDICE F – ACURÁCIAS DA ESTRATÉGIA MST-CLU-DS COM
CLASSIFICADORES SVM, K-NN, RF E NB**

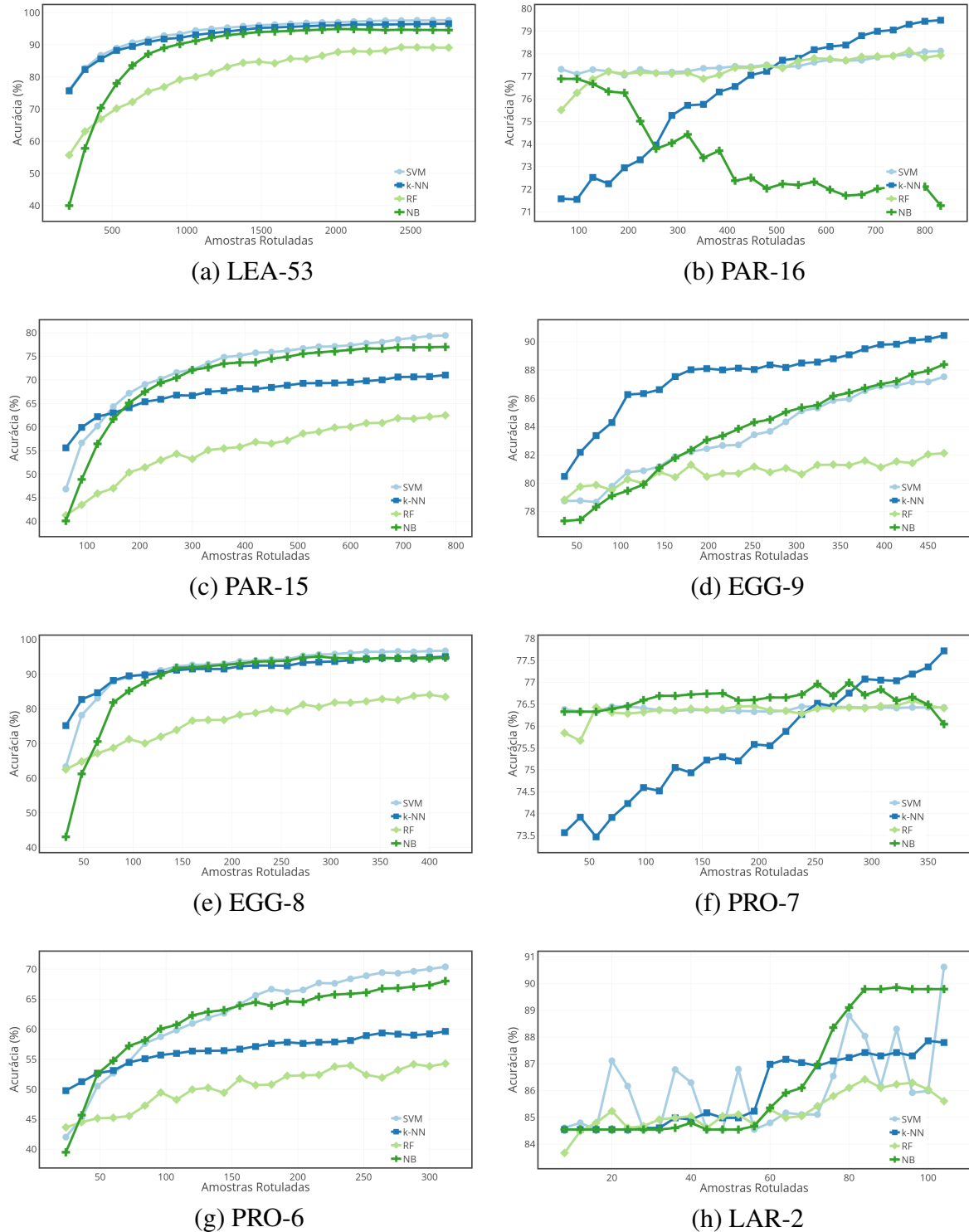
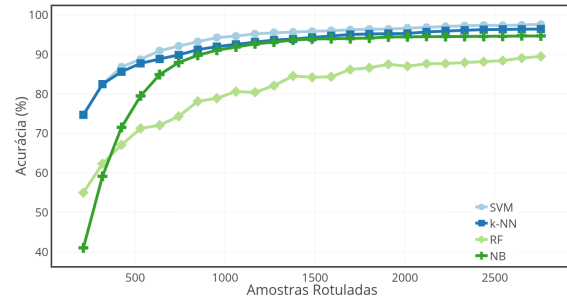
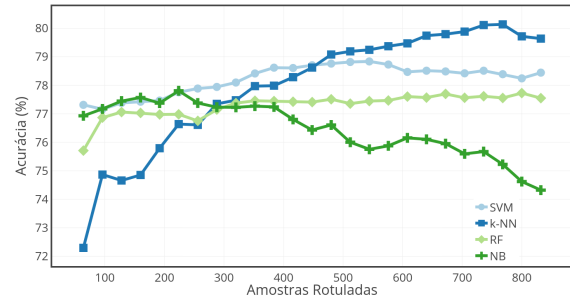


Figura 45: Médias das acurácias obtidas pelas estratégia de aprendizado ativo MST-CLU-DS com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

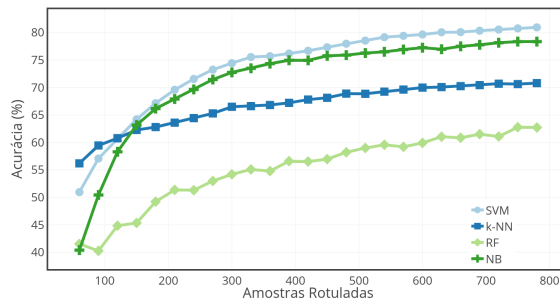
**APÊNDICE G – ACURÁCIAS DA ESTRATÉGIA MST-CLU-RDS COM
CLASSIFICADORES SVM, K-NN, RF E NB**



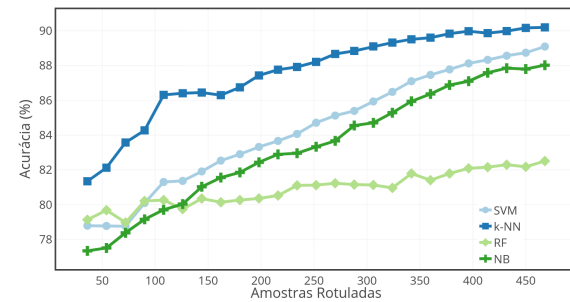
(a) LEA-53



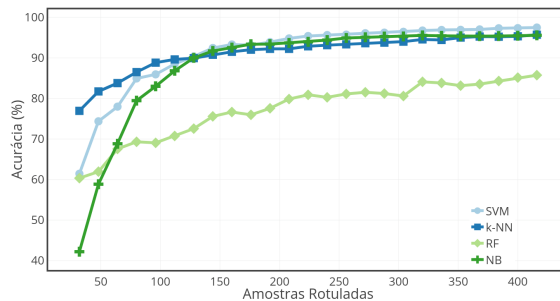
(b) PAR-16



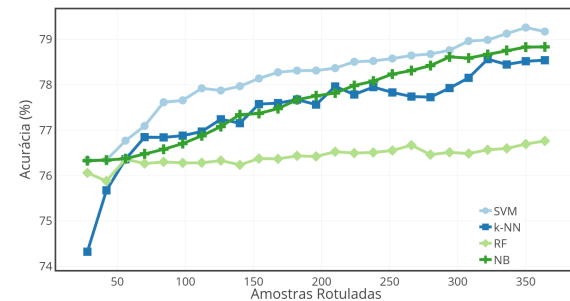
(c) PAR-15



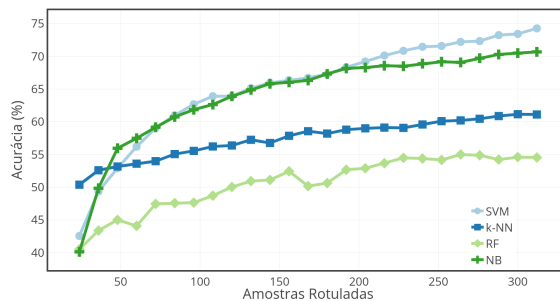
(d) EGG-9



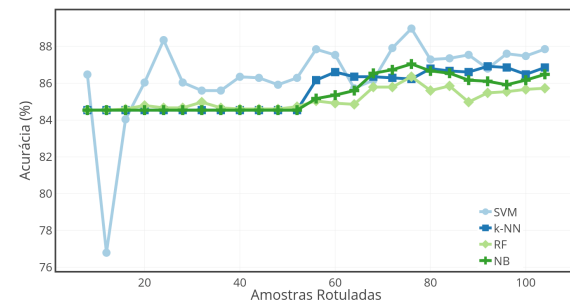
(e) EGG-8



(f) PRO-7



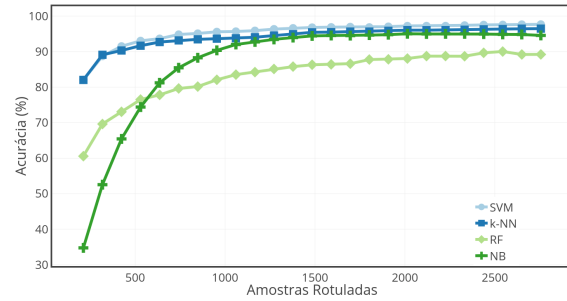
(g) PRO-6



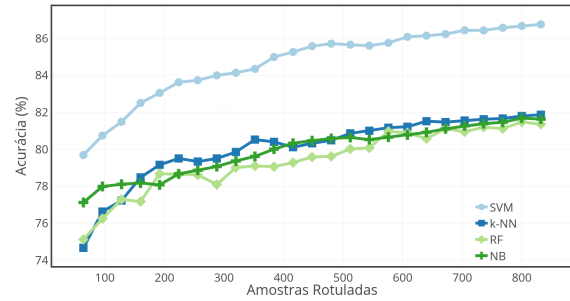
(h) LAR-2

Figura 46: Médias das acurácias obtidas pela estratégia de aprendizado ativo MST-CLU-RDS com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

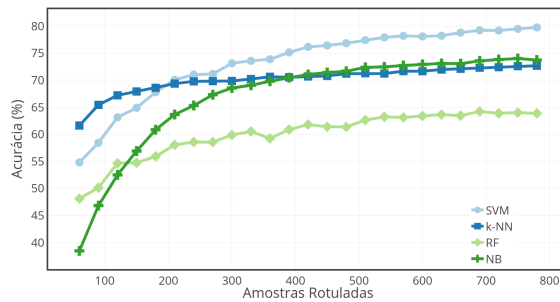
**APÊNDICE H – ACURÁCIAS DA ESTRATÉGIA RDS COM CLASSIFICADORES
SVM, K-NN, RF E NB**



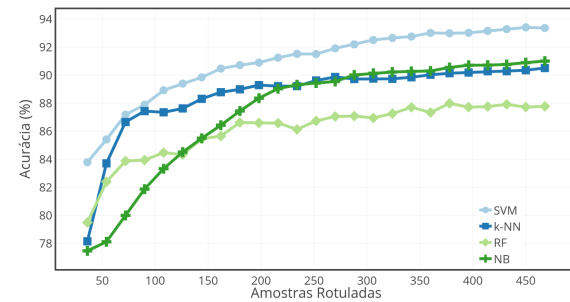
(a) LEA-53



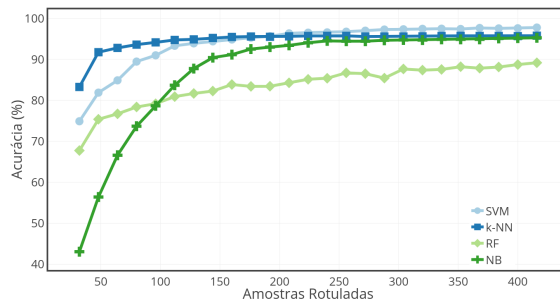
(b) PAR-16



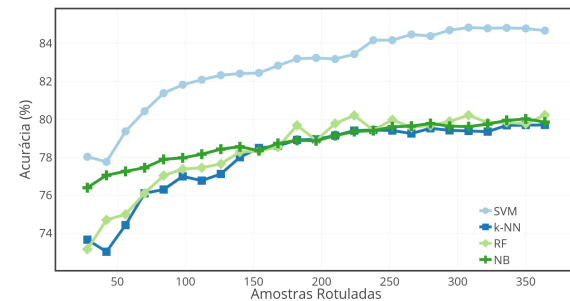
(c) PAR-15



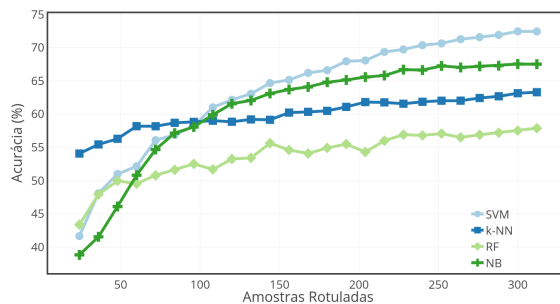
(d) EGG-9



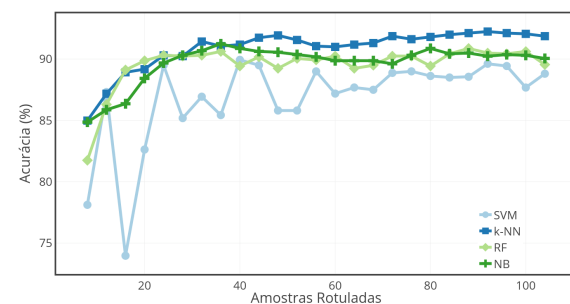
(e) EGG-8



(f) PRO-7



(g) PRO-6



(h) LAR-2

Figura 47: Médias das acurácias obtidas pelas estratégia de aprendizado ativo RDS com os classificadores SVM, k -NN, RF e NB nos conjuntos (a) LEA-53, (b) PAR-16, (c) PAR-15, (d) EGG-9, (e) EGG-8, (f) PRO-7, (g) PRO-6 e (h) LAR-2.

**APÊNDICE I – MÉDIA EM SEGUNDOS DOS TEMPOS DE ORGANIZAÇÃO,
SELEÇÃO E CLASSIFICAÇÃO OBTIDOS PELAS ESTRATÉGIAS
MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS E RDS COM
O CLASSIFICADOR SVM EM CADA CONJUNTO DE DADOS**

Tabela 13: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto LEA-53.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	1.692e+00	3.546e-01	2.809e+00
MST-CLU-DDE	1.735e+00	6.056e-02	2.752e+00
MST-CLU-DS	1.752e+00	1.929e-05	2.761e+00
MST-CLU-RDS	1.694e+00	2.998e-01	2.703e+00
RDS	1.646e+00	3.149e-01	2.900e+00

Tabela 14: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PAR-16.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	1.987e+01	2.362e+00	2.483e-01
MST-CLU-DDE	3.297e+01	3.715e-01	2.425e-01
MST-CLU-DS	3.300e+01	1.626e-04	2.221e-01
MST-CLU-RDS	3.121e+01	1.669e+00	2.366e-01
RDS	1.993e+01	1.985e+00	2.857e-01

Tabela 15: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PAR-15.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	3.247e+00	5.238e-01	2.301e-01
MST-CLU-DDE	4.126e+00	8.446e-03	2.520e-01
MST-CLU-DS	4.238e+00	4.041e-05	2.553e-01
MST-CLU-RDS	3.894e+00	4.563e-01	2.521e-01
RDS	3.273e+00	4.606e-01	2.221e-01

Tabela 16: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto EGG-9.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	2.779e+00	3.034e-01	9.670e-02
MST-CLU-DDE	4.104e+00	6.273e-02	9.683e-02
MST-CLU-DS	4.267e+00	4.024e-05	9.194e-02
MST-CLU-RDS	3.948e+00	2.446e-01	8.973e-02
RDS	2.801e+00	2.857e-01	1.096e-01

Tabela 17: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto EGG-8.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	4.348e-01	5.593e-02	7.162e-02
MST-CLU-DDE	5.564e-01	4.503e-03	7.097e-02
MST-CLU-DS	5.606e-01	1.588e-05	7.311e-02
MST-CLU-RDS	5.082e-01	4.881e-02	6.971e-02
RDS	4.417e-01	5.250e-02	7.768e-02

Tabela 18: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PRO-6.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	1.646e+00	1.425e-01	3.344e-02
MST-CLU-DDE	2.604e+00	2.758e-02	3.520e-02
MST-CLU-DS	2.588e+00	2.972e-05	3.668e-02
MST-CLU-RDS	2.438e+00	1.341e-01	3.568e-02
RDS	1.642e+00	1.327e-01	3.003e-02

Tabela 19: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto PRO-7.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	8.066e+00	7.179e-01	4.700e-02
MST-CLU-DDE	2.318e+01	1.750e-01	4.001e-02
MST-CLU-DS	2.322e+01	1.132e-04	2.178e-02
MST-CLU-RDS	2.201e+01	3.575e-01	2.713e-02
RDS	8.066e+00	4.781e-01	3.419e-02

Tabela 20: Média em segundos dos tempos de organização, seleção e classificação obtidos pelas estratégias MS-CLU, MST-CLU-DDE, MST-CLU-DS, MST-CLU-RDS e RDS com o classificador SVM no conjunto LAR-2.

Estratégia	Tempo Organização	Tempo Seleção	Tempo Classificação
MS-CLU	2.432e+00	2.310e-02	6.431e-03
MST-CLU-DDE	1.658e+00	6.077e-02	4.775e-03
MST-CLU-DS	4.284e+00	1.515e-05	6.930e-03
MST-CLU-RDS	1.154e+00	1.815e-02	4.219e-03
RDS	2.299e+00	3.476e-02	7.075e-03