



**UNIVERSIDADE DO ESTADO DO RIO GRANDE DO NORTE
UNIVERSIDADE FEDERAL RURAL DO SEMI-ÁRIDO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA
COMPUTAÇÃO**



HIURY NOGUEIRA DE ARAÚJO

**UTILIZANDO APRENDIZADO SEMISSUPERVISIONADO
MULTIDSCRIÇÃO EM PROBLEMAS DE
CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO**

MOSSORÓ – RN
2017

HIURY NOGUEIRA DE ARAÚJO

**UTILIZANDO APRENDIZADO SEMISSUPERVISIONADO
MULTIDESCRICÃO EM PROBLEMAS DE
CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação – associação ampla entre a Universidade do Estado do Rio Grande do Norte e a Universidade Federal Rural do Semi-Árido, para a obtenção do título de Mestre em Ciência da Computação.

Orientador: ARAKEN DE MEDEIROS SANTOS, Prof.
Dr.

MOSSORÓ – RN
2017

© Todos os direitos estão reservados a Universidade Federal Rural do Semi-Árido. O conteúdo desta obra é de inteira responsabilidade do (a) autor (a), sendo o mesmo, passível de sanções administrativas ou penais, caso sejam infringidas as leis que regulamentam a Propriedade Intelectual, respectivamente, Patentes: Lei n° 9.279/1996 e Direitos Autorais: Lei n° 9.610/1998. O conteúdo desta obra tomar-se-á de domínio público após a data de defesa e homologação da sua respectiva ata. A mesma poderá servir de base literária para novas pesquisas, desde que a obra e seu (a) respectivo (a) autor (a) sejam devidamente citados e mencionados os seus créditos bibliográficos.

A658u Araújo, Hiury Nogueira de.
UTILIZANDO APRENDIZADO SEMISSUPERVISIONADO
MULTIDESCRICÃO EM PROBLEMAS DE CLASSIFICAÇÃO
HIERÁRQUICA MULTIRRÓTULO / Hiury Nogueira de
Araújo. - 2017.
126 f. : il.

Orientador: Araken de Medeiros Santos.
Dissertação (Mestrado) - Universidade Federal
Rural do Semi-árido, Programa de Pós-graduação em
Ciência da Computação, 2017.

1. Aprendizado Semissupervisionado. 2.
Classificação Hierárquica Multirrótulo. 3. Co-
training. 4. Self-training. I. Santos, Araken de
Medeiros, orient. II. Título.

O serviço de Geração Automática de Ficha Catalográfica para Trabalhos de Conclusão de Curso (TCC's) foi desenvolvido pelo Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (USP) e gentilmente cedido para o Sistema de Bibliotecas da Universidade Federal Rural do Semi-Árido (SISBI-UFERSA), sendo customizado pela Superintendência de Tecnologia da Informação e Comunicação (SUTIC) sob orientação dos bibliotecários da instituição para ser adaptado às necessidades dos alunos dos Cursos de Graduação e Programas de Pós-Graduação da Universidade.

AGRADECIMENTOS

Agradeço a toda minha família, pois todos têm a sua participação para que eu pudesse chegar onde estou agora.

A minha mãe Sônia, meu irmão Geraldo e meu padrasto Nailson, por terem me dado as oportunidades que me deram, pelo apoio e motivação que sempre foram presentes.

A minha tia Jeanne e seu marido Denys, que desde a minha chegada à Mossoró sempre me ajudaram de diversos modos, praticamente me acolheram dentro de sua residência. Gosto muito de ambos.

A minha linda vó Maria de Jesus Nogueira, que a tenho também como mãe, pois desde criança me ensinou os valores que carrego até hoje. Infelizmente ela não se encontra entre nós, mas a sua lembrança sempre estará presente, não só em mim, mas em toda minha família.

Ao Eudes, pois foi a pessoa que me apresentou, há muitos anos, a computação, a qual me apaixonei logo no primeiro contato e decidi que iria levá-la como profissão por toda minha vida.

Aos meus amigos de Russas, que sempre me apoiaram, respeitaram minha ausência por diversos momentos e continuaram sendo meus amigos, mesmo com a distância.

Aos professores e técnicos do PPgCC, que sempre estavam disponíveis quando precisei dos mesmos.

Ao meu orientador, Araken, a quem possuo uma grande admiração e gratidão, pois é uma pessoa muito profissional, dedicada, sabendo brincar e cobrar nas horas certas. Foi um prazer termos trabalhado juntos. Muito obrigado por tudo.

Enfim, obrigado a todos que de alguma forma contribuíram para este momento.

Acredite em si próprio e chegará um dia em que os outros não terão outra escolha senão acreditar com você.

Cynthia Kersey

RESUMO

A classificação de dados é uma tarefa aplicada em diversas áreas do conhecimento, sendo assim, foco de constantes pesquisas. A classificação de dados pode ser dividida de acordo com a disposição dos dados, sendo estes rotulados ou não rotulados. Uma abordagem vem se mostrando bastante eficiente ao se trabalhar com conjuntos de dados contendo dados rotulados e não rotulados, esta chamada de aprendizado semissupervisionado, seu objetivo é classificar os dados não rotulados através da quantidade de dados rotulados contidos no conjunto, melhorando sua taxa de acerto. Tais dados podem ser classificados com mais de um rótulo, conhecida como classificação multirrótulo. Além disso, estes dados podem estar organizados de forma hierárquica, contendo assim, uma relação entre os mesmos, esta, por sua vez, denominada classificação hierárquica. Neste trabalho é proposto a utilização do aprendizado semissupervisionado multidescrição, que é uma das vertentes do aprendizado semissupervisionado, em problemas de classificação hierárquica multirrótulo, com o objetivo de investigar se o aprendizado semissupervisionado é uma abordagem apropriada para resolver o problema de baixa dimensionalidade de dados. Uma análise experimental dos métodos verificou que o aprendizado supervisionado obteve melhor desempenho contra as abordagens semissupervisionadas, contudo, o aprendizado semissupervisionado pode vir a ser uma abordagem amplamente utilizada, pois, há bastante o que ser contribuído nesta área.

Palavras-chave: Aprendizado Semissupervisionado; Classificação Hierárquica Multirrótulo; *Co-training*; *Self-training*.

ABSTRACT

Data classification is a task applied in various areas of knowledge, therefore, the focus of ongoing research. Data classification can be divided according to the available data, which are labeled or not labeled. One approach has proven very effective when working with data sets containing labeled and unlabeled data, this called semi-supervised learning, your objective is to label the unlabeled data by using the amount of labeled data in the data set, improving their success rate. Such data can be classified with more than one label, known as multi-label classification. Furthermore, these data can be organized hierarchically, thus containing a relation therebetween, this called hierarchical classification. This work proposes the use of multi-view semi-supervised learning, which is one of the semissupervisionado learning aspects, in problems of hierarchical multi-label classification, with the objective of investigating whether semi-supervised learning is an appropriate approach to solve the problem of low dimensionality of data. An experimental analysis of the methods found that supervised learning had a better performance than semi-supervised approaches, however, semi-supervised learning may be a widely used approach, because, there is plenty to be contributed in this area.

Keywords: Semi-supervised Learning; Hierarchical Multi-label Classification; Co-training; Self-training.

LISTA DE FIGURAS

Figura 2.1 –	Categorias do aprendizado de máquina de acordo com o grau de supervisão.	26
Figura 2.2 –	Problemas de classificação.	31
Figura 2.3 –	Processo utilizando LP	33
Figura 2.4 –	Processo utilizado pelo BR.	34
Figura 2.5 –	Processo de treinamento e predição do CC.	35
Figura 2.6 –	Estrutura em Árvore	37
Figura 2.7 –	Estrutura em DAG	37
Figura 2.8 –	Transformação do problema hierárquico em problema de classificação plana	39
Figura 2.9 –	Classificação local	40
Figura 2.10 –	Abordagem <i>Top-Down</i>	41
Figura 2.11 –	Abordagem <i>Big-Bang</i> (Global)	42
Figura 4.1 –	Equação da fórmula matemática da métrica de <i>Hamming Loss</i>	51
Figura 4.2 –	Equação da fórmula matemática da métrica de <i>Subset Accuracy</i>	52
Figura 4.3 –	Equação da fórmula matemática da métrica de <i>Accuracy</i>	52
Figura 4.4 –	Equação da fórmula matemática da métrica de <i>One-Error</i>	53
Figura 4.5 –	Equação da fórmula matemática da métrica de <i>Average Precision</i>	53
Figura 4.6 –	Equação da fórmula matemática da métrica de <i>Ranking Loss</i>	53
Figura 4.7 –	Equação da fórmula matemática da métrica de <i>Hierarchical Loss</i>	54
Figura 4.8 –	Árvore da Configuração dos Experimentos Semissupervisionados	55
Figura 5.1 –	Métodos de Classificação Hierárquica Multirrótulo Supervisionados no Nível 1 da Hierarquia.	61
Figura 5.2 –	Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Visão Simples no Nível 1 da Hierarquia	61
Figura 5.3 –	Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Multivisão no Nível 1 da Hierarquia.	63
Figura 5.4 –	Método Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados no Nível 1 da Hierarquia.	63
Figura 5.5 –	Método Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados no Nível 1 da Hierarquia.	64
Figura 5.6 –	Método Semissupervisionado Visão Simples Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 1 da Hierarquia	65
Figura 5.7 –	Método Semissupervisionado Multivisão Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 1 da Hierarquia	65
Figura 5.8 –	Métodos Semissupervisionado Multivisão <i>versus</i> Semissupervisionado Visão Simples <i>versus</i> Supervisionado no Nível 1 da Hierarquia	66
Figura 5.9 –	Métodos de Classificação Hierárquica Multirrótulo Supervisionados no Nível 2 da Hierarquia.	67

Figura 5.10 – Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Visão Simples no Nível 2 da Hierarquia	68
Figura 5.11 – Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Multivisão no Nível 2 da Hierarquia.	69
Figura 5.12 – Método Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados no Nível 2 da Hierarquia. . . .	70
Figura 5.13 – Método Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados no Nível 2 da Hierarquia.	71
Figura 5.14 – Método Semissupervisionado Visão Simples Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 2 da Hierarquia	71
Figura 5.15 – Método Semissupervisionado Multivisão Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 2 da Hierarquia	72
Figura 5.16 – Métodos Semissupervisionado Multivisão <i>versus</i> Semissupervisionado Visão Simples <i>versus</i> Supervisionado no Nível 2 da Hierarquia	73
Figura 5.17 – Métodos de Classificação Hierárquica Multirrótulo Supervisionados no Nível 3 da Hierarquia.	74
Figura 5.18 – Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Visão Simples no Nível 3 da Hierarquia	75
Figura 5.19 – Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Multivisão no Nível 3 da Hierarquia.	77
Figura 5.20 – Método Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados no Nível 3 da Hierarquia. . . .	77
Figura 5.21 – Método Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados no Nível 3 da Hierarquia.	78
Figura 5.22 – Método Semissupervisionado Visão Simples Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 3 da Hierarquia	79
Figura 5.23 – Método Semissupervisionado Multivisão Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 3 da Hierarquia	79
Figura 5.24 – Métodos Semissupervisionado Multivisão <i>versus</i> Semissupervisionado Visão Simples <i>versus</i> Supervisionado no Nível 3 da Hierarquia.	80
Figura 5.25 – Desempenho Entre Níveis no Aprendizado Supervisionado, Semissupervisionado <i>Self-Training</i> e <i>Co-Training</i>	82

LISTA DE TABELAS

Tabela 2.1 –	Conjunto de exemplos no formato atributo-valor.	30
Tabela 2.2 –	Exemplo de saída utilizando o método LP.	33
Tabela 2.3 –	Exemplo de saída ao utilizar o método Relevância Binária.	34
Tabela 4.1 –	Resumo das Bases de Dados Hierárquica Multirrótulo	49
Tabela A.1 –	Média e o desvio padrão dos métodos BR, CC e LP supervisionados Nível 1	97
Tabela A.2 –	Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado visão simples Nível 1.	98
Tabela A.3 –	Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado multivisão Nível 1	99
Tabela A.4 –	Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado visão simples Nível 1.	100
Tabela A.5 –	Média e o desvio padrão dos percentuais iniciais de exemplos rotulados para o método semissupervisionado multivisão Nível 1	101
Tabela A.6 –	Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado visão simples Nível 1.	102
Tabela A.7 –	Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado multivisão Nível 1.	103
Tabela A.8 –	Média e desvio padrão das abordagens semissupervisionada multivisão <i>versus</i> semissupervisionada visão simples <i>versus</i> abordagem supervisionada Nível 1.	104
Tabela A.9 –	Média e o desvio padrão dos métodos BR, CC e LP supervisionados Nível 2	105
Tabela A.10 –	Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado visão simples Nível 2.	106
Tabela A.11 –	Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado multivisão Nível 2	107
Tabela A.12 –	Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado visão simples Nível 2.	108
Tabela A.13 –	Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado multivisão Nível 2	109
Tabela A.14 –	Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado visão simples Nível 2.	110
Tabela A.15 –	Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado multivisão Nível 2.	111
Tabela A.16 –	Média e desvio padrão das abordagens semissupervisionada multivisão <i>versus</i> semissupervisionada visão simples <i>versus</i> abordagem supervisionada Nível 2.	112

Tabela A.17 –	Média e o desvio padrão dos métodos BR, CC e LP supervisionados Nível 3	113
Tabela A.18 –	Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado visão simples Nível 3.	114
Tabela A.19 –	Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado multivisão Nível 3	115
Tabela A.20 –	Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado visão simples Nível 3.	116
Tabela A.21 –	Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado multivisão Nível 3	117
Tabela A.22 –	Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado visão simples Nível 3.	118
Tabela A.23 –	Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado multivisão Nível 3.	119
Tabela A.24 –	Média e desvio padrão das abordagens semissupervisionada multivisão <i>versus</i> semissupervisionada visão simples <i>versus</i> abordagem supervisionada Nível 3.	120
Tabela A.25 –	Média dos Melhores Resultados do Aprendizado Supervisionado nos Três Níveis da Classificação Hierárquica Multirrótulo.	121
Tabela A.26 –	Média dos Melhores Resultados do Aprendizado Semissupervisionado Visão Simples nos Três Níveis da Classificação Hierárquica Multirrótulo	123
Tabela A.27 –	Média dos Melhores Resultados do Aprendizado Semissupervisionado Multivisão nos Três Níveis da Classificação Hierárquica Multirrótulo	125

LISTA DE SIGLAS

1-Err	<i>One-Error</i>
Acc	<i>Accuracy</i>
AM	Aprendizado de Máquina
AvPrec	<i>Average Precision</i>
BR	<i>Binary Relevance</i>
HiLoss	<i>Hierarchical Loss</i>
HLoss	<i>Hamming Loss</i>
HMC-LMLP	<i>Hierarchical Multi-Label Classification with Local Multi-Layer Perceptron</i>
HMC-SSBR	<i>Hierarchical Multilabel Classification using Semi-Supervised Binary Relevance</i>
HMC-SSLP	<i>Hierarchical Multilabel Classification using Semi-Supervised Label Powerset</i>
HMC-SSRAkEL	<i>Hierarchical Multilabel Classification using Semi-Supervised Random k-Labelsets</i>
HMC-SSRAkELd	<i>Hierarchical Multilabel Classification using Semi-Supervised Random k disjoint Labelsets</i>
HOMER	<i>Hierarchy Of Multilabel classiFieRS</i>
IA	Inteligência Artificial
JRip	<i>Extended Repeated Incremental Pruning</i>
kNN	<i>k-Nearest Neighbor</i>
LP	<i>Label Powerset</i>

MIPS	<i>Munich Information Center for Protein Sequences</i>
MLP	<i>Multi-Layer Perceptron</i>
PCT	<i>Predictive Clustering Trees</i>
RLoss	<i>Ranking Loss</i>
SAcc	<i>Subset Accuracy</i>
WEKA	<i>Waikato Environment for Knowledge Analysis</i>

SUMÁRIO

1 INTRODUÇÃO	17
1.1 MOTIVAÇÃO	19
1.2 OBJETIVO.....	20
1.3 ESTRUTURA DO TRABALHO	20
2 APRENDIZADO DE MÁQUINA	22
2.1 CONCEITOS BÁSICOS.....	22
2.2 TIPOS DE APRENDIZADO	25
2.3 APRENDIZADO SEMISSUPERVISIONADO	26
2.3.1 Descrição Simples	28
2.3.2 Multidescrição.....	28
2.4 CLASSIFICAÇÃO DE DADOS.....	29
2.4.1 Classificação Tradicional	30
2.4.2 Classificação Multirrótulo.....	31
2.4.2.1 <i>Label Powerset</i>	32
2.4.2.2 Relevância Binária	33
2.4.2.3 <i>Classifier Chains</i>	35
2.4.3 Classificação de Dados Hierárquico	36
2.4.3.1 Transformação do Problema Hierárquico em Problema de Classificação Plana.....	38
2.4.3.2 Predição Hierárquica com Algoritmos de Classificação Plana.....	39
2.4.3.3 Classificação Hierárquica <i>Top-Down</i>	40
2.4.3.4 Classificação Hierárquica <i>Big-Bang</i>	41
2.5 CONSIDERAÇÕES FINAIS	42
3 TRABALHOS RELACIONADOS	43
3.1 APRENDIZADO SEMISSUPERVISIONADO	43
3.2 CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO.....	45
3.3 APRENDIZADO SEMISSUPERVISIONADO EM PROBLEMAS DE CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO	46
3.4 CONSIDERAÇÕES FINAIS	47
4 MATERIAIS E MÉTODOS	48
4.1 BASES DE DADOS.....	48

4.1.2 Bases de Dados Hierárquica Multirrótulo.....	48
4.1.2.1 <i>Cellcycle</i>	49
4.1.2.2 <i>Church</i>	49
4.1.2.3 <i>Derisi</i>	50
4.1.2.4 <i>Eisen</i>	50
4.1.2.5 <i>Gasch1</i>	50
4.1.2.6 <i>Gasch2</i>	50
4.1.2.7 SPO.....	50
4.2 MÉTRICAS DE AVALIAÇÃO.....	50
4.2.1 Métricas Baseadas em Bipartição.....	51
4.2.1.1 <i>Hamming Loss</i>	51
4.2.1.2 <i>Subset Accuracy</i>	52
4.2.1.6 <i>Accuracy</i>	52
4.2.2 Métricas Baseadas em <i>Ranking</i>	52
4.2.2.1 <i>One-Error</i>	52
4.2.2.2 <i>Average Precision</i>	53
4.2.2.6 <i>Ranking Loss</i>	53
4.2.3 <i>Hierarchical Loss</i>	54
4.3 CONFIGURAÇÃO DOS EXPERIMENTOS.....	54
4.4 TESTES ESTATÍSTICOS.....	55
4.5 CONSIDERAÇÕES FINAIS.....	57
5 RESULTADOS DOS EXPERIMENTOS.....	58
5.1 MÉTODOS HIERÁRQUICOS MULTIRRÓTULO: PRIMEIRO NÍVEL.....	60
5.1.1 Melhores Resultados dos Métodos BR, LP e CC no Contexto Supervisionado.....	60
5.1.2 Melhores Resultados dos Métodos BR, LP e CC no Contexto Semissupervisionado Visão Simples.....	61
5.1.3 Melhores Resultados dos Métodos BR, LP e CC no Contexto Semissupervisionado Multivisão.....	62
5.1.4 Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados.....	63
5.1.5 Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados.....	63

5.1.6 Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados	64
5.1.7 Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados	65
5.1.8 Abordagem Semissupervisionada Multivisão <i>versus</i> Semissupervisionada Visão Simples <i>versus</i> Supervisionada.....	66
5.2 MÉTODOS HIERÁRQUICOS MULTIRRÓTULO: SEGUNDO NÍVEL.....	66
5.2.1 Melhores Resultados dos Métodos BR, LP e CC no Contexto Supervisionado.....	67
5.2.2 Melhores Resultados dos Métodos BR, LP e CC no Contexto Semissupervisionado Visão Simples	68
5.2.3 Melhores Resultados dos Métodos BR, LP e CC no Contexto Semissupervisionado Multivisão	69
5.2.4 Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados	70
5.2.5 Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados	70
5.2.6 Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados	71
5.2.7 Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados	72
5.2.8 Abordagem Semissupervisionada Multivisão <i>versus</i> Semissupervisionada Visão Simples <i>versus</i> Supervisionada.....	72
5.3 MÉTODOS HIERÁRQUICOS MULTIRRÓTULO: TERCEIRO NÍVEL	73
5.3.1 Melhores Resultados dos Métodos BR, LP e CC no Contexto Supervisionado.....	74
5.3.2 Melhores Resultados dos Métodos BR, LP e CC no Contexto Semissupervisionado Visão Simples	75
5.3.3 Melhores Resultados dos Métodos BR, LP e CC no Contexto Semissupervisionado Multivisão	76
5.3.4 Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados	77
5.3.5 Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados	77

5.3.6 Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados	78
5.3.7 Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados	79
5.3.8 Abordagem Semissupervisionada Multivisão <i>versus</i> Semissupervisionada Visão Simples <i>versus</i> Supervisionada.....	80
5.4 DESEMPENHO ENTRE NÍVEIS	81
5.4.1 Desempenho do Aprendizado Supervisionado, Semissupervisionado <i>Self-Training</i> e <i>Co-Training</i>	81
5.5 CONSIDERAÇÕES FINAIS	82
6 CONCLUSÃO.....	85
6.1 TRABALHOS FUTUROS.....	87
REFERÊNCIAS	88
APÊNDICE A - VALORES DOS EXPERIEMNTOS REALIZADOS.....	93

Capítulo 1

INTRODUÇÃO

Aprendizado de Máquina (AM) é uma área dedicada ao estudo e desenvolvimento de técnicas computacionais que permitam o aperfeiçoamento do desempenho em alguma tarefa através de experiências acumuladas. Essas experiências são passadas através de conjuntos de dados, estes são apresentados aos algoritmos, que através do princípio de indução, realizam uma generalização do problema, conseguindo, após esta etapa, solucionar novos problemas apresentados (MITCHELL, 1997).

Os algoritmos (técnicas) que realizam essa tarefa, são divididos em três grupos, mas inicialmente eram divididos em apenas dois: o aprendizado supervisionado, o qual possui o objetivo de realizar a classificação de novos exemplos após a fase de treinamento, e o não-supervisionado, que sua meta é explorar ou descrever um conjunto de dados. O terceiro, denominado aprendizado semissupervisionado, surgiu para tratar um problema que comumente era encontrado: conjuntos de dados com informações rotuladas e não rotuladas. O aprendizado supervisionado trabalha apenas com dados rotulados, enquanto o aprendizado não-supervisionado com não rotulados. Por sua vez, o semissupervisionado trabalha com um conjunto de dados possuindo tanto dados rotulados quanto não rotulados, aprendendo, assim, com os dados rotulados do conjunto de dados e rotula os não rotulados deste mesmo conjunto, objetivando conseguir um classificador com melhor taxa de acerto (acurácia).

Ademais, o aprendizado semissupervisionado pode ser dividido em dois grupos, descrição simples e multidescrição, onde este primeiro utiliza todo o conjunto de entrada para gerar o classificador, gerando uma única visão (descrição simples) do problema, e o segundo divide este conjunto de entrada em duas visões (multidescrição), gerando dois

classificadores. O aprendizado semissupervisionado e o aprendizado supervisionado têm como objetivo a realização da classificação de dados.

Tradicionalmente, na classificação de dados um exemplo do conjunto de dados está relacionado a apenas um rótulo, ou a uma classe. Esta tarefa é conhecida como classificação simples, tradicional, unirrótulo ou monorrótulo. Mas existem um a série de domínios onde os exemplos estão associados a um conjunto de rótulos, por exemplo, em classificação de textos, imagens e na área da bioinformática. Nesse contexto, a tarefa de classificação é chamada multirrótulo (TSOUMAKAS; KATAKIS; VLAHAVAS, 2010).

Além disso, a grande maioria dos trabalhos propõem soluções onde o conjunto de classes de um problema estão dispostos em um mesmo nível, sendo conhecida como classificação plana. Neste tipo de classificação, a relação hierárquica entre as classes é completamente desconsiderada, onde se supõe independência entre as mesmas.

Por outro lado, existe um outro tipo de classificação em que a relação hierárquica entre as classes é considerada, esta é chamada de classificação hierárquica. Existe um grande número de problemas complexos que utilizam essa abordagem. Nesses problemas uma ou mais classes podem ser divididas em subclasses ou agrupadas em superclasses (RODRIGUES, 2014).

Contudo, recentemente, tem-se observado que muitos trabalhos estão abordando problemas em que a classificação hierárquica e multirrótulo estão sendo tratadas conjuntamente (CERRI, 2010; CERRI; CARVALHO, 2009; CERRI; BARROS; CARVALHO, 2011). Estes problemas são conhecidos como problemas de classificação hierárquica multirrótulo. E em outros já estão inserindo neste contexto o aprendizado semissupervisionado. Mas neste caso, apenas a aprendizagem semissupervisionada descrição simples está sendo abordada (SANTOS, 2012; SANTOS; CANUTO, 2014).

Neste contexto, considerando as vantagens do aprendizado semissupervisionado multidescrição e o aumento no número de aplicações hierárquica multirrótulo, este trabalho propõe um estudo da aplicação de técnicas de aprendizado semissupervisionado multidescrição em tarefas de classificação hierárquica multirrótulo em conjuntos de dados biológicos, mais especificamente, em funções de proteínas, que são macromoléculas formadas por longas sequências de aminoácidos e que executam quase todas as funções

celulares nos seres vivos, de modo a identificar seu desempenho em comparação com outras abordagens de aprendizado.

1.1. MOTIVAÇÃO

Como dito anteriormente, a principal tarefa do aprendizado supervisionado e semissupervisionado é a classificação de dados, estes podem ser encontradas em diversas áreas de domínios, como categorização de textos, imagens, áudios, vídeos, dados biológicos e médicos. Tal tarefa vem sendo estudada há muitos anos, tendo como frutos, inúmeros estudos científicas que têm contribuído para que a classificação de dados esteja evoluindo para uma melhor rotulação de dados. Ademais, outras abordagens na classificação de dados surgiram com o passar do tempo, como por exemplo, a classificação multirrótulo e a hierárquica multirrótulo, ambas já introduzidas anteriormente e que abrem ainda mais as possibilidades de novos estudos, sejam utilizando cada técnica separadamente ou utilizando-as em conjunto.

É muito fácil encontrar trabalhos dos quais combinam o aprendizado supervisionado com a classificação multirrótulo ou com a classificação hierárquica, mas quando passamos a aplicar o aprendizado semissupervisionado na classificação multirrótulo ou hierárquica, as contribuições encontradas já diminuem consideravelmente, e quando combinado o semissupervisionado com a classificação hierárquica multirrótulo, esta é muito rara, tendo em vista, que até o presente momento, foi-se utilizado apenas a técnica de descrição simples na combinação destas abordagens.

Um dos principais problemas na classificação de dados é a baixa taxa de acerto, oriunda de uma quantidade pequena de dados rotulados, ou seja, baixa dimensionalidade de exemplos de dados. Uma técnica que parte de uma pequena quantidade de dados rotulados e aumenta este número de dados rotulados, é o aprendizado semissupervisionado.

Contudo, o problema citado anteriormente e a possível solução com o aprendizado semissupervisionado tem motivado o autor à realizar contribuições nesta área, tendo em vista que é uma área que tem muito a ser estudado, sendo este trabalho o carro chefe para suas contribuições.

A pesquisa desenvolvida para este trabalho de dissertação, realizou um estudo no qual foi investigado o problema de baixa dimensionalidade de dados em problemas de classificação hierárquica multirrótulo utilizando o aprendizado semissupervisionado multidescrição, visando detectar o desempenho desta em relação a outras abordagens, salientando que este é um estudo exclusivo, até o presente momento, contribuindo, assim, com a comunidade científica.

1.2. OBJETIVOS

O objetivo do trabalho proposto é o estudo da aplicação de técnicas de aprendizado semissupervisionado multidescrição em tarefas de classificação hierárquica multirrótulo, afim de verificar se a utilização desta traz melhorias nos resultados do processo de classificação de dados. Para tal, alguns objetivos secundários são descritos a seguir:

1. Implementar técnicas de aprendizado semissupervisionado, classificação hierárquica, classificação multirrótulo e os classificadores base, para serem utilizadas em conjunto nas bases de dados e gerar os resultados para este estudo;
2. Avaliar os resultados por meio de diferentes medidas específicas para classificação hierárquica multirrótulo;

1.3. ESTRUTURA DO TRABALHO

O restante deste trabalho está organizado da seguinte forma:

- No Capítulo 2: Descreve a fundamentação teórica para o desenvolvimento deste trabalho, dando ênfase nas abordagens específicas utilizadas;
- O Capítulo 3: Apresenta o estado da arte, os trabalhos relacionados aos temas abordados neste trabalho;
- No Capítulo 4: Aborda os procedimentos adotados para a realização dos experimentos, as bases de dados utilizadas e as medidas de avaliação dos resultados;

- No Capítulo 5: São descritos os resultados dos experimentos;
- O Capítulo 6: Apresenta as considerações finais;
- O Apêndice A: São apresentados os experimentos realizados.

Capítulo 2

APRENDIZADO DE MÁQUINA

Aprendizado de Máquina é uma área interdisciplinar, a mesma surgiu para sanar a necessidade de resolver os problemas computacionais cada vez mais complexos e mais presentes no dia a dia e trabalhar com grandes quantidades de dados, tarefas difíceis ou até impossíveis de serem realizadas por seres humanos. Ou seja, AM é uma área cujo objetivo é o desenvolvimento de técnicas computacionais que tornem viáveis ao computador tomar decisões baseadas em experiências acumuladas, ou seja, tratar da questão da construção de algoritmos capazes de adquirir conhecimento de forma automática, melhorando seu desempenho na tomada de decisões por meio de soluções bem-sucedidas de problemas anteriores (MITCHELL, 1997; MONARD, 2003). A aquisição de conhecimento automática só é possível através da aplicação do princípio da inferência chamado de indução.

A indução é uma forma de raciocínio lógico. Tal conceito permite que conclusões genéricas sejam obtidas a partir de um conjunto particular de exemplos. O raciocínio indutivo origina-se em conceito específico e generalizado, ou seja, da parte para o todo (MONARD, 2003). O processo de aquisição de conhecimento baseado em indução é conhecido como aprendizado indutivo.

2.1. CONCEITOS BÁSICOS

Na literatura, encontra-se uma variedade de algoritmos de aprendizado de máquina, cada uma possui suas características que lhes são peculiares. Porém, existem alguns termos que podem ser aplicados à grande parte desses métodos.

Santos (2012) apresenta os principais termos associados à área de aprendizado de máquina:

- **Atributo:** um atributo descreve uma característica ou aspecto de um exemplo. Normalmente, esses atributos podem ser de dois tipos: discretos ou contínuos. Atributos discretos, também chamados de atributos nominais ou categóricos, contêm um número finito de valores. Por sua vez, os atributos contínuos podem assumir um número infinito de valores. Um importante ponto a ser considerado na escolha dos atributos que serão utilizados para descrever os exemplos, é a sua capacidade preditiva, ou seja, a capacidade de discriminação entre os exemplos (MONARD, 2003; FACELLI et al., 2011).
- **Classe ou rótulo:** uma classe ou rótulo, também chamado de atributo-alvo, descreve o conceito-meta, ou seja, o conceito que se deseja aprender para tornar viável a realização de previsões a seu respeito.
- **Exemplo:** um exemplo, também chamado de padrão ou instância, descreve um objeto de um determinado conceito que se deseja aprender através de um vetor de valores de características ou atributos (MONARD, 2003). Um exemplo pode conter, além do vetor de características, um atributo especial, indicando a classe a que este exemplo pertence.
- **Conjunto de exemplos:** um conjunto de exemplos, também chamado de conjunto de dados, é composto por um número de exemplos com seus respectivos valores de atributos.
- **Conjunto de treinamento:** utilizado como entrada pelos algoritmos de aprendizado. A partir deste conjunto são construídos (induzidos) modelos ou hipóteses. Assim, devem possuir uma quantidade representativa da população dos dados do domínio.
- **Conjunto de teste:** utilizado para avaliar o modelo construído. Idealmente, o conjunto de teste não deve ter exemplos em comum com o conjunto de treinamento.

- Conjunto de validação: utilizado, em alguns casos, para realizar ajustes no modelo construído pelo algoritmo de aprendizado. Os exemplos desse conjunto não são utilizados diretamente na construção dos modelos, mas são utilizados para o seu ajuste. Dessa maneira, esses exemplos são indiretamente "vistos" durante o processo de aprendizado, obrigando que os exemplos de validação sejam distintos dos exemplos de teste.
- Indutor: algoritmo de aprendizado que utiliza um processo indutivo para gerar a sua hipótese ou modelo. Essa hipótese é utilizada para classificar uma instância ou exemplo.
- Classificador ou hipótese: dado um conjunto de exemplos de treinamento, um indutor ou algoritmo de aprendizado gera como saída um classificador (ou hipótese) de forma que, dado um novo padrão, ele possa prever sua classe com a maior precisão possível (MONARD, 2003).
- *Overfitting* (superajustamento): o *overfitting* ocorre quando, após o treinamento, o classificador gerado é muito específico para o conjunto de exemplos de treinamento utilizado. Assim, a maioria das instâncias que não pertence ao conjunto de exemplos utilizado durante o processo de aprendizado é classificada erroneamente. Desse modo, pode-se dizer que, durante o treinamento, o classificador gerado foi superajustado ao conjunto de treinamento, incapacitando-o a classificar ou generalizar corretamente exemplos diferentes daqueles utilizados na fase de treinamento.
- *Underfitting*: o *underfitting* consiste no caso inverso ao *overfitting*, ou seja, o classificador gerado ajustou-se muito pouco ao conjunto de treinamento, incapacitando-o a classificar corretamente tanto os exemplos do conjunto de dados de treinamento, quanto os exemplos em um conjunto de teste (MONARD, 2003).
- Acurácia: a acurácia de um classificador é uma medida de desempenho obtida por este classificador para uma determinada

tarefa. Essa medida é calculada de acordo com a taxa de predições corretas (precisão) ou incorretas (taxa de erro) realizadas por esse classificador para um determinado conjunto de dados. Desse modo, a acurácia pode ser interpretada como a quantidade de padrões que esse classificador acertou ou errou de acordo com a classe a que este padrão de fato pertence. Essa taxa de acurácia é, em geral, estimada utilizando um conjunto diferente do conjunto utilizado no processo de aprendizado, chamado de conjunto de teste.

2.2. TIPOS DE APRENDIZADO

O aprendizado indutivo pode ser dividido em duas categorias principais: aprendizado supervisionado e aprendizado não-supervisionado (MONARD, 2003). Na literatura podem ser encontrados como aprendizado preditivo e aprendizado descritivo, respectivamente. A diferença entre ambos diz respeito à forma como é realizado o processo de generalização do conhecimento.

Aprendizado Supervisionado: A meta é encontrar uma função (também chamada de modelo ou hipótese) a partir dos dados de treinamento que possa ser utilizada para prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores de seus atributos de entrada. Para isso, cada objeto do conjunto de treinamento deve possuir atributos de entrada e saída (FACELLI et al., 2011). Para os atributos de saída do tipo discreto, o problema é denominado de classificação e para os do tipo contínuo, é chamado de regressão.

Após a criação do modelo (processo de treinamento), quando um novo exemplo, não-rotulado, for apresentado ao tal, o mesmo deverá rotulá-lo, através das experiências anteriores, com uma das possíveis classes determinadas pelo conjunto de treinamento.

Aprendizado Não-Supervisionado: A meta é explorar ou descrever um conjunto de dados. Os algoritmos de AM utilizados nessas tarefas não fazem uso do atributo de saída (FACELLI et al., 2011). O aprendizado não-supervisionado possui três métodos para tratamento do problema. Facelli et al. (2011) descreve os métodos da seguinte forma:

- Agrupamento: Os dados são agrupados de acordo com sua similaridade.
- Sumarização: O objetivo é encontrar uma descrição simples e compacta para o conjunto de dados.
- Associação: Consiste em encontrar padrões frequentes de associações entre atributos de um conjunto de dados.

Há alguns anos, a comunidade científica vem estudando uma terceira categoria de aprendizado, surgida através da junção do aprendizado supervisionado com o aprendizado não-supervisionado, como podemos visualizar na Figura 2.1. A essa nova categoria de aprendizado dá-se o nome de aprendizado semissupervisionado (MATSUBARA; MONARD; BATISTA, 2005). A mesma será descrita na próxima sessão.

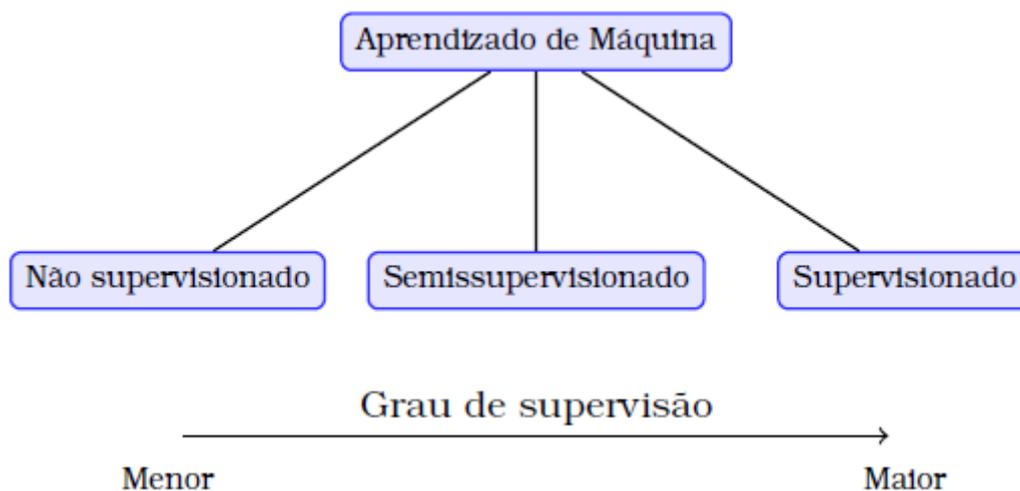


Figura 2.1: Categorias do aprendizado de máquina de acordo com o grau de supervisão

Fonte: Metz, 2011

2.3. APRENDIZADO SEMISSUPERVISIONADO

Em situações nas quais exemplos rotulados são escassos em comparação à quantidade de exemplos não rotulados, pode ser interessante procurar por algoritmos de aprendizado que explorem esses dois “mundos” conjuntamente (BRAGA, 2010). Esta é a abordagem aplicada nos algoritmos de aprendizado semissupervisionado. Para este, é apresentado um conjunto de dados contendo uma quantidade de dados rotulados L (*Labeled*), que geralmente são quantidades muito menores, em relação ao conjunto de

dados não rotulados U (*Unlabeled*). Induzir classificadores a partir de um conjunto de dados com essas características, leva à indução de classificadores com uma baixa acurácia. Um dos principais objetivos do aprendizado semissupervisionado é rotular mais exemplos e, desse modo, incrementar a quantidade de exemplos do conjunto de dados de treinamento. Assim, pode-se apresentar um algoritmo supervisionado, para que este possa induzir classificadores mais precisos.

Santos (2012) cita duas vantagens do aprendizado semissupervisionado:

1. O potencial de reduzir a necessidade de uma grande quantidade de dados rotulados, em domínios nos quais somente um pequeno conjunto de padrões rotulados estão disponíveis.
2. Quando o especialista não apresenta um total conhecimento sobre o conceito a ser aprendido, ou seja, esse especialista tem apenas o conhecimento de alguns padrões de um determinado conjunto de dados, apresentando, assim, grande dificuldade de rotular exemplos para incrementar o conjunto de dados de treinamento.

Como exemplos de técnicas de aprendizado semissupervisionado bastante utilizados podemos citar o Self-Training (ROSENBERG; HEBERT; SCHNEIDERMAN, 2005) e o Co-Training (BLUM; MITCHELL, 1998). Na literatura, os mesmos podem ser conhecidos por algoritmo de descrição simples e multidescrição, respectivamente. Estes serão descritos na próxima seção.

Esses algoritmos usam a abordagem de *Bootstrapping*. Abney (2008) explica esta, como uma abordagem para o aprendizado semissupervisionado cuja ideia principal é obter um classificador inicial usando um algoritmo base e exemplos rotulados. Após isso, o classificador inicial rotula novos exemplos ainda não rotulados. Os exemplos rotulados pelo classificador inicial e que possuem maior confiança são retirados do conjunto de exemplos não rotulados e inseridos no conjunto de exemplos rotulados. Esse processo se repete até que todos os exemplos do conjunto de exemplos não rotulados tenham sido rotulados ou algum critério não permita mais a rotulação.

Esta abordagem está sujeita à propagação de erros, pois caso um exemplo não rotulado seja rotulado erradamente e inserido no conjunto de treinamento para as futuras

novas rotulações, este irá propagar o erro durante todo o processo, comprometendo assim a rotulação dos exemplos não rotulados.

2.3.1. Descrição Simples

O *Self-Training* é um algoritmo de aprendizado semissupervisionado, este é o que mais se destaca quando a característica de descrição simples é abordada. Esta característica se dá ao fato deste gerar apenas um classificador utilizando todos os atributos de entrada do conjunto de dados apresentados para a fase treinamento, por isso denominado de descrição simples ou visão simples, como abordado em outras bibliografias.

O processo de geração do seu primeiro classificador, inicia com a realização do treinamento a partir dos dados rotulados disponíveis no conjunto de dados. Este primeiro classificador gerado, será o responsável por rotular os exemplos não rotulados. Após a realização deste, serão selecionados, dos novos exemplos rotulados, os que possuem uma maior taxa de confiança e estes serão inseridos no conjunto de treinamento e os demais retornarão para o conjunto de não rotulados. O próximo passo é gerar um novo classificador, com o conjunto de treinamento reformulado e rotular os não rotulados novamente. Todo o processo se repete até que todos os exemplos não rotulados sejam rotulados ou não, obedecendo um determinado critério de parada, este definido no início do processo. Vale salientar que a quantidade de dados não rotulados iniciais e a escolha da quantidade de melhores exemplos a serem selecionados por iteração, são critérios determinados antes do início do processo.

2.3.2. Multidescrição

Ao contrário da descrição simples, a abordagem multidescrição trata o problema gerando dois classificadores. O algoritmo *Co-Training* é muito utilizado, se tratando dessa abordagem, esta é abordada como multidescrição ou multivisão, ambas as formas encontradas em diversas bibliografias.

Nesse algoritmo, dois classificadores são inicialmente gerados, cada um utilizando uma descrição diferente dos exemplos rotulados, ou seja, repartindo os atributos de entrada, dos exemplos rotulados, em dois grupos. Uma vez que são utilizadas

duas descrições, pode-se utilizar o mesmo algoritmo de aprendizado supervisionado para induzir as duas hipóteses (classificadores) ou dois algoritmos de aprendizado supervisionado distintos. Desse modo, é gerado um classificador sobre os mesmos exemplos, mas cada um induzido seguindo visões (descrições) diferentes. Cada visão é suficiente, está apta, para prever as classes perfeitamente e que as duas visões são condicionalmente independentes (DU; LING; ZHOU, 2010). Assim, pode-se afirmar que o *Co-Training* se baseia na ideia de que cada um dos classificadores rotulará exemplos que conterão informações desconhecidas pelo outro classificador.

Uma vez gerados os dois classificadores, os exemplos não-rotulados são apresentados aos dois classificadores, os quais terão suas saídas combinadas. Em seguida, os exemplos que apresentarem o grau de confiança de classificação mais altos são inseridos no conjunto de exemplos rotulados. A partir desse conjunto incrementado de exemplos rotulados, são gerados dois novos classificadores e o processo é repetido até que algum critério de parada, estabelecido antes do início do processo, seja satisfeito. Assim como no *Self-Training*, critérios de quantidade de dados não rotulados iniciais e dos melhores exemplos a serem inseridos no conjunto de rotulados também devem ser definidos antes de iniciar o processo.

2.4. CLASSIFICAÇÃO DE DADOS

A classificação de dados é um processo no qual, a partir de um conjunto de dados brutos, são extraídas informações por meio de categorização. Desse modo, um processo de classificação consiste na atribuição de rótulos aos dados, de tal forma que esses rótulos confirmem informações aos dados categorizados sob o mesmo rótulo (COSTA, 2008).

O conceito de classificação de dados é bastante comum em várias áreas de conhecimento, tais como: Computação, Engenharia, Matemática, Estatística, dentre outras. Na computação, uma área de estudo com amplo interesse em classificação de dados é a área de aprendizado de máquina, mais especificamente o aprendizado supervisionado. Este tem como objetivo, já citado anteriormente, prever um rótulo ou valor que caracterize um novo exemplo, com base nos valores de seus atributos de entrada.

Para que este objetivo seja alcançado, esses atributos de entrada devem possuir um rótulo, ou seja, um atributo que os classifiquem como pertencentes à uma determinada classe. Este atributo pode ser encontrado, na literatura, com as seguintes nomenclaturas: atributo alvo, atributo meta, classe ou rótulo. Assim é formado o conjunto de treinamento, que será apresentado ao algoritmo de aprendizado supervisionado, e que por fim gerará o classificador.

Ainda neste contexto, a classificação de dados pode ser dividida em classificação tradicional ou unirrótulo (do inglês, *single-label classification*) e classificação multirrótulo (do inglês, *multi-label classification*). A classificação tradicional considera que para cada exemplo é atribuído um único rótulo. Já na classificação multirrótulo, um único exemplo pode pertencer a uma ou mais classes.

Continuando neste contexto, temos a classificação hierárquica, onde a classificação dos dados pode estar hierarquicamente organizada.

Essas formas de classificação serão descritas nas próximas sessões.

2.4.1. Classificação Tradicional

Na classificação tradicional, cada exemplo do conjunto de dados possui apenas um único rótulo. Na Tabela 2.1, vemos que para N exemplos E_1, \dots, E_N . Cada exemplo rotulado E_i assume a forma (x_i, y_i) , onde $x_i = (x_{i1}, x_{i2}, \dots, x_{iR}) \in X$ e y_i é o valor do atributo classe C . Ou seja, na classificação tradicional, cada exemplo é associado a uma única classe (rótulo) de um conjunto de classes C , de modo que $|C| > 1$. Essa tarefa é chamada de classificação unirrótulo. Se $|C| = 2$, então a tarefa de aprendizado é chamada de classificação unirrótulo binária. Por outro lado, se $|C| > 2$, a tarefa é chamada de classificação unirrótulo multiclasse.

Tabela 2.1: Conjunto de exemplos no formato atributo-valor.

	X_1	X_2	\dots	X_R	C
E_1	X_{11}	X_{12}	\dots	X_{1R}	y_1
E_2	X_{21}	X_{22}	\dots	X_{2R}	y_2
\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
E_N	X_{N1}	X_{N2}	\dots	X_{NR}	y_N

2.4.2. Classificação Multirrótulo

Vimos que na classificação tradicional, cada exemplo do conjunto de dados está relacionado a apenas uma classe (rótulo). Entretanto, existe um grupo de problemas reais de classificação, conhecidas como problemas de classificação multirrótulo, em que cada exemplo pode pertencer simultaneamente a mais de uma classe (FACELLI *et al.*, 2011).

Nos últimos anos, tem-se observado um crescimento contínuo no número de publicações relacionadas à classificação multirrótulo, indicando um grande interesse da comunidade científica nessa área de pesquisa (METZ, 2011).

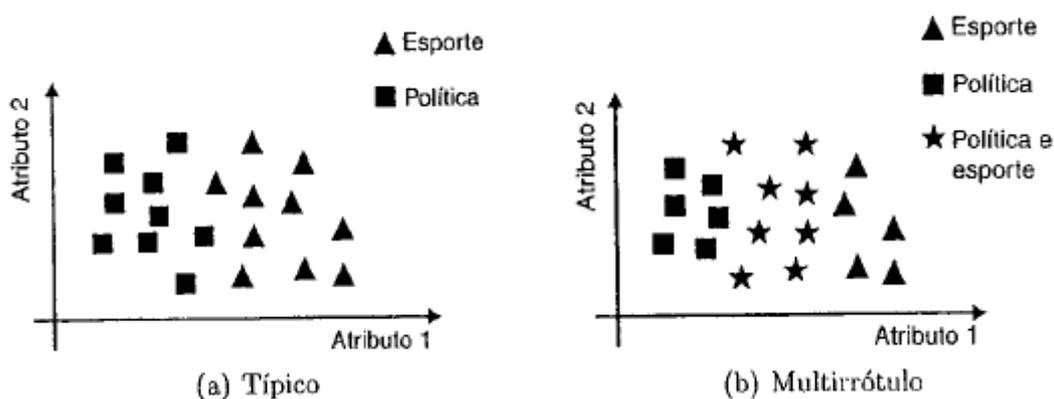


Figura 2.2: Problemas de classificação

Fonte: Facelli *et al.*, 2011

Na Figura 2.2, é apresentada uma comparação entre um problema de classificação tradicional (unirrótulo) e um problema de classificação multirrótulo. A Figura 2.2.a ilustra um problema de classificação típica (unirrótulo) no qual um exemplo pode pertencer à classe "Política" ou à classe "Esporte", mas nunca às duas classes simultaneamente. Por outro lado, a Figura 2.2.b ilustra um exemplo de classificação multirrótulo, em que os exemplos podem pertencer simultaneamente às classes "Política" e "Esporte", podendo, então, serem classificados como pertencentes à classe "Política e Esporte".

Na literatura, podem ser encontrados diversos métodos e técnicas para o tratamento de problemas de classificação multirrótulo. Estes sendo divididos em duas abordagens: Independente de algoritmo e Dependente de algoritmo.

Em Facelli *et al.* (2011), essas duas abordagens são descritas da seguinte forma:

- Independente de Algoritmo: Utiliza algoritmos tradicionais de classificação para tratar problemas multirrótulo, transformando o problema multirrótulo original em um conjunto de problemas simples-rótulo.
- Dependente de Algoritmo: Cria algoritmos específicos para tratar o problema multirrótulo. Esses algoritmos podem ser baseados em técnicas de classificação convencionais, como máquina de vetores de suporte (SVMs, do inglês, *Support Vectors Machine*) e árvores de decisão, ou podem ser especificamente desenvolvidos para a classificação multirrótulo.

Neste trabalho foi utilizada a abordagem Independente de Algoritmo. Das técnicas utilizadas nessa abordagem, destacam-se os métodos *Label Powerset* – LP, *Binary Relevance* – BR e *Classifier Chain* – CC.

2.4.2.1. *Label Powerset*

O método de transformação de problema menos comum, mas ainda assim bastante utilizado é o *Label Powerset* (KATAKIS; TSOUMAKAS; VLAHAVAS, 2008). Neste método, cada subconjunto diferente de rótulos de L é considerado como uma única classe da nova tarefa de classificação unirrótulo. Desse modo, um classificador unirrótulo $C: X \rightarrow P(L)$ é treinado, onde $P(L)$ é o powerset de L , contendo todos os subconjuntos de rótulos possíveis. Assim, dada uma nova instância, o classificador unirrótulo C retorna como saída a classe mais provável, que neste caso é um conjunto de rótulos. A Tabela 2.2 mostra o exemplo de saída usando o método *Label Powerset*. A Figura 2.3 ilustra o processo utilizado pelo LP.

Uma das vantagens deste método é que, no LP, as correlações entre os rótulos são consideradas (SANTOS, 2012). Contudo, é suscetível ao fato de, no caso de haver um número muito grande de subconjuntos de rótulos, o número de rótulos de uma classe pode crescer exponencialmente, resultando em muitas classes com poucos exemplos associados, aumentando o custo computacional do LP e diminuindo a acurácia dos classificadores. Além disso, o LP só pode prever confiavelmente conjuntos de rótulos

(*labelsets*) observados no conjunto de treinamento. Esta é uma importante limitação, uma vez que novos *labelsets* tipicamente aparecerão em instâncias de teste.

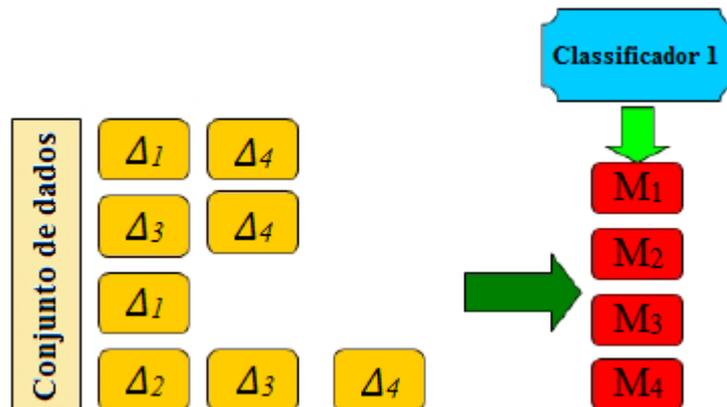


Figura 3: Processo utilizando LP

Fonte: Adaptado de Santos, 2012

Tabela 2.2: Exemplo de saída utilizando o método LP

Exemplo	Rótulo
1	$\{\Delta_{1,4}\}$
2	$\{\Delta_{3,4}\}$
3	$\{\Delta_1\}$
4	$\{\Delta_{2,3,4}\}$

2.4.2.1. Relevância Binária

Dentre os métodos de transformação do problema, o mais popular é o método de Relevância Binária (do inglês, *Binary Relevance*) (TSOUMAKAS; KATAKIS, 2007). Neste método, o problema multirrotulado é dividido em $|L|$ problemas de classificação unirrótulo binário, onde $|L|$ é a quantidade de rótulos contidos em L . Assim, pode-se afirmar que a predição de cada rótulo é considerada como uma tarefa independente.

No método de Relevância Binária, o conjunto de dados original é transformado em $|L|$ conjuntos de dados D_{Δ_j} , cada um dos quais contendo todos os exemplos do conjunto de dados original, rotulados como Δ_j se o conjunto de rótulos do exemplo original contém Δ_j e como: $\sim\Delta_j$, caso contrário. Com isso, um classificador binário $C_{\Delta_j}: X \rightarrow \{\sim\Delta_j, \Delta_j\}$ é treinado para cada diferente rótulo $\Delta_j \in L$. Assim, uma vez que todos os classificadores tenham sido treinados, para a classificação de uma nova instância é retornada como saída

a união dos Δ_j rótulos que são positivamente preditos pelos $|L|$ classificadores. Na Tabela 2.3, são apresentados os conjuntos de dados produzidos ao utilizar o método de Relevância Binária. A Figura 2.4 ilustra o processo utilizado pelo BR.

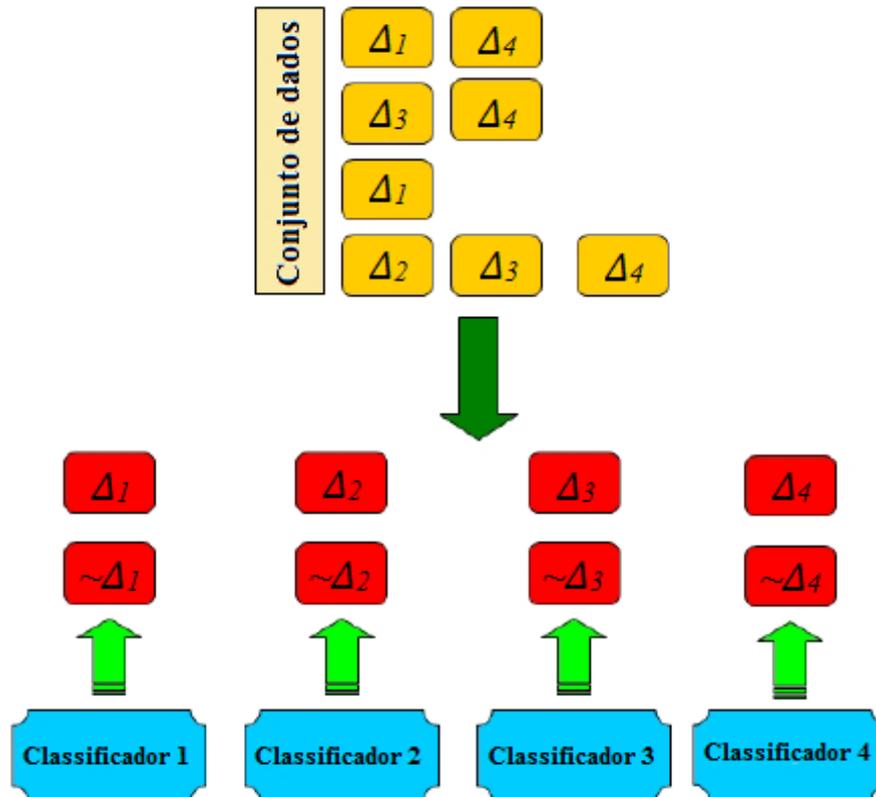


Figura 2.4: Processo utilizado pelo BR

Fonte: Adaptado de Santos, 2012

Uma das principais críticas ao BR é o fato de neste método não serem consideradas as correlações entre os rótulos, descartando assim informações que podem ser importantes à tarefa de classificação (SANTOS, 2012). Desse modo, o modelo gerado pelo método de Relevância Binária pode levar à uma pouca capacidade de generalização. Além disso, o modelo de classificação gerado é de baixa legibilidade, uma vez que é gerado um modelo em separado para cada um dos rótulos associado a um exemplo de teste.

Tabela 2.3: Exemplo de saída ao utilizar o método Relevância Binária

Exemplo	Rótulos	Exemplo	Rótulos	Exemplo	Rótulos	Exemplo	Rótulos
1	$\{\Delta_1\}$	1	$\{\sim\Delta_2\}$	1	$\{\sim\Delta_3\}$	1	$\{\Delta_4\}$
2	$\{\sim\Delta_1\}$	2	$\{\sim\Delta_2\}$	2	$\{\Delta_3\}$	2	$\{\Delta_4\}$
3	$\{\Delta_1\}$	3	$\{\sim\Delta_2\}$	3	$\{\sim\Delta_3\}$	3	$\{\sim\Delta_4\}$
4	$\{\sim\Delta_1\}$	4	$\{\Delta_2\}$	4	$\{\Delta_3\}$	4	$\{\Delta_4\}$

2.4.2.1. Classifier Chain

Em Read *et al.* (2011), foi introduzido um novo classificador multirrótulo, denominado *Classifier Chain*. Este método surgiu para solucionar uma desvantagem do método BR, que era a não consideração da dependência dos rótulos. Para isso, o CC também utiliza $|L|$ classificadores binários, assim como o BR, mas os classificadores são ligados em uma corrente (*chain*), utilizando uma ordem definida aleatoriamente, na qual cada classificador lida com um problema binário associado a um rótulo $Y_j \in L$ (CHERMAN, 2013).

Cada classificador binário gerado contém seus atributos de entrada acrescidos dos classificadores binários gerados nas etapas anteriores, ou seja, a predição de todos os classificadores construídos anteriormente são utilizados como entrada para construção do próximo metaclassificador (METZ, 2011). Esse processo é repetido até que tenham sido criados n classificadores binários, um para cada monorrótulo em L . O processo de treinamento e predição é mostrado na Figura 2.5.

A vantagem deste método é manter a simplicidade e eficiência do BR e considerar a correlação entre os rótulos para aumentar sua capacidade preditiva.

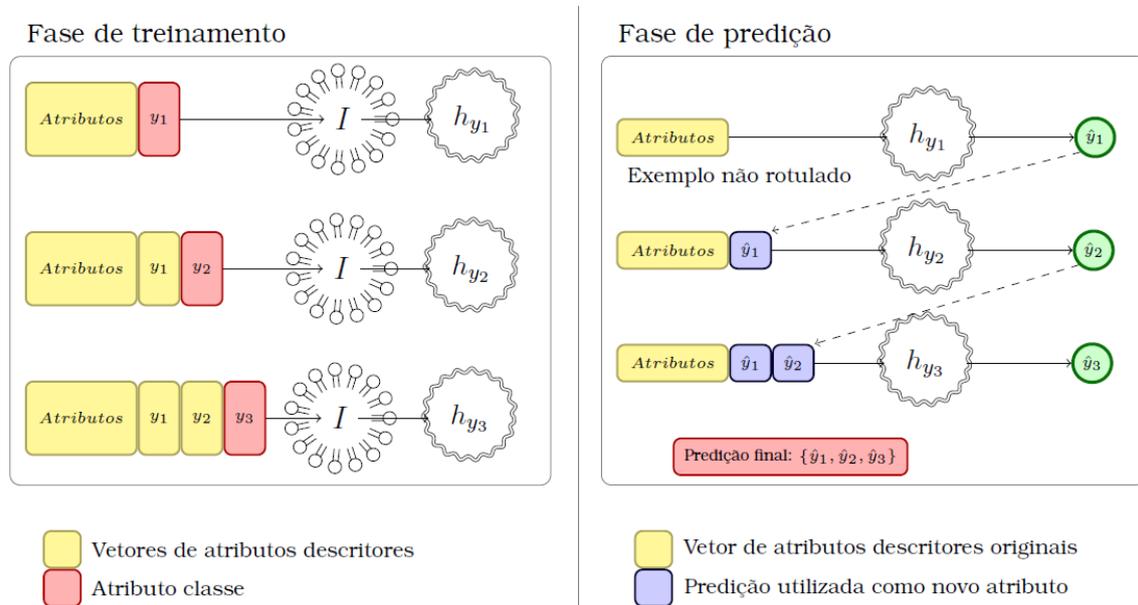


Figura 2.5: Processo de treinamento e predição do CC

Fonte: Adaptado de Metz, 2011

2.4.3. Classificação Hierárquica

A grande parte das pesquisas relacionadas com a Aprendizagem de Máquina estão relacionadas com problemas de classificação plana, onde a classificação unirrótulo ou multirrótulo pode ser utilizada para rotular os exemplos, entretanto, neste tipo de classificação, não são consideradas as relações entre classes. No entanto, existe uma área de estudo onde a relação entre classes é estudada, esta é conhecida como classificação hierárquica.

Nos problemas de classificação hierárquica, as classes podem apresentar uma relação de taxonomia ou dependência, formando subclasses e superclasses. A classificação plana, em que não se tem uma relação hierárquica entre as classes, é o tipo mais comum de classificação. Contudo, se a hierarquia das classes é conhecida ou pode ser construída, sua consideração pode levar à indução de classificadores com maior acurácia preditiva (FACELLI *et al.*, 2011).

Na classificação plana o único ponto de conexão entre as classes envolvidas no problema é o nó raiz, já na classificação hierárquica, as classes que se ramificam a partir do nó raiz podem se relacionar com outras classes, chamadas de classes descendentes ou subclasses.

O tipo de hierarquia empregada na representação dos relacionamentos entre as classes e o nível hierárquico em que as predições são realizadas, são os dois aspectos que merecem maior destaque entre os aspectos que caracterizam um problema de classificação hierárquica.

Os tipos de estruturas empregadas em problemas de classificação hierárquica são: árvore e grafo acíclico direcionado (DAG, do inglês, *Directed Acyclic Graph*). A principal diferença entre a estrutura de árvore e a estrutura DAG é que, na estrutura de árvore, cada, nó, exceto o nó raiz, possui somente um nó pai, enquanto no DAG, cada nó, exceto o nó raiz, pode ter um ou mais nós pai. Em ambos os esquemas de classificação, plana e hierárquica, os nós representam as classes do problema e o nó raiz corresponde a “qualquer classe”, denotado pela ausência total de conhecimento sobre a classe de um objeto. As estruturas de árvore e de DAG são ilustradas nas Figuras 2.6 e 2.7, respectivamente.

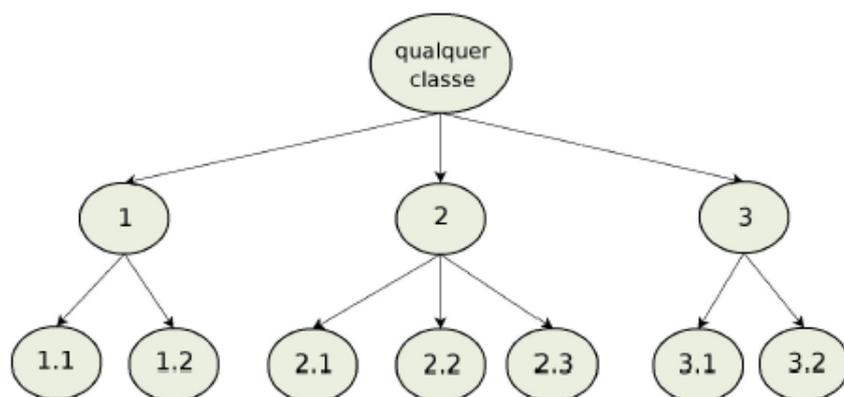


Figura 2.6: Estrutura em Árvore

Fonte: Freitas; Carvalho, 2007

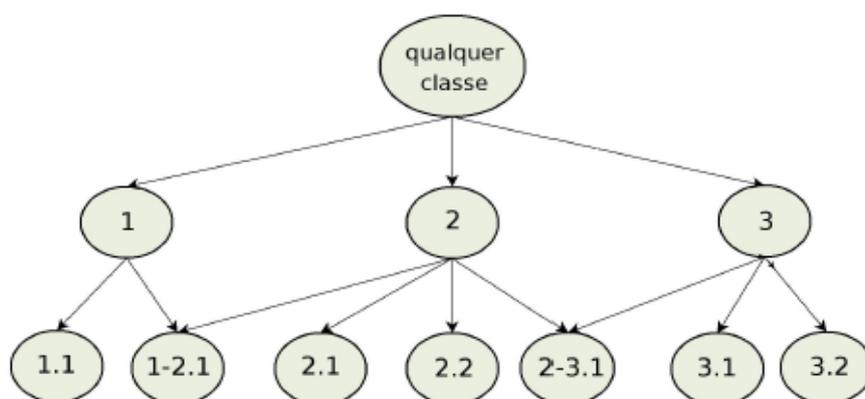


Figura 2.7: Estrutura em DAG

Fonte: Freitas; Carvalho, 2007

Problemas de classificação hierárquica frequentemente têm como objetivo a classificação de uma nova entrada em um dos nós folhas. Segundo Facelli *et al.* (2011), a profundidade de predição do algoritmo está relacionada a ele sempre realizar predições em nós folhas (predição mandatária em nós folha) ou a ele ser capaz de realizar predições em qualquer nível da hierarquia, incluindo as folhas (predição não mandatária em nós folha). Embora a confiabilidade da predição possa ser maximizada na predição não mandatária em nós folha, a classificação torna-se menos específica e pode assim ter menos utilidade.

Em Santos (2012), é feita uma análise entre as estruturas de classe de árvore e a estrutura DAG:

- Em estruturas de classes em árvore, a dificuldade de predição é diretamente proporcional à profundidade de um nível na

hierarquia, de modo que, quanto mais profundo é o nível na hierarquia, mais difícil é o processo de predição. Tal afirmação se deve ao fato de que, neste tipo de estrutura, as classes nos níveis mais profundos na hierarquia representam informações mais específicas e são produzidas por modelos induzidos a partir de um menor número de instâncias que seu ancestral. Por essa razão, estas classes são mais difíceis de prever corretamente.

- Em estruturas DAG, a análise é mais complexa. Como um nó filho pode ter mais de um nó pai, alguns modelos de classificação em níveis mais profundos podem ter sido induzidos a partir de um número maior de instâncias que seus ancestrais.

Na prática, mesmo para DAGs, a taxa de acurácia de predição diminui com o aumento no nível da classe (FREITAS; CARVALHO, 2007).

Ainda em Freitas e Carvalho (2007), quatro tipos de abordagens para lidar com problemas hierárquicos podem ser citados: transformação do problema hierárquico em um problema de classificação plana, predição hierárquica com algoritmos de classificação plana, classificação *top-down* e classificação *big-bang*.

A seguir, será feita uma breve descrição destes quatro tipos de esquema de classificação hierárquica.

2.4.3.1. Transformação do Problema Hierárquico em Problema de Classificação Plana

Esta abordagem consiste na transformação do problema hierárquico em problema de classificação plana, com isso, os algoritmos de classificação plana são utilizados para a resolução do problema.

Na criação do classificador, é ignorada a hierarquia de classes, apenas as classes dos nós folhas são consideradas, como ilustra a Figura 2.8. Com isso, se uma classe de um nó folha é atribuída a um exemplo, as classes ancestrais dispostas hierarquicamente também são atribuídas a essa classe, fornecendo assim uma solução indireta para o problema de classificação hierárquica.

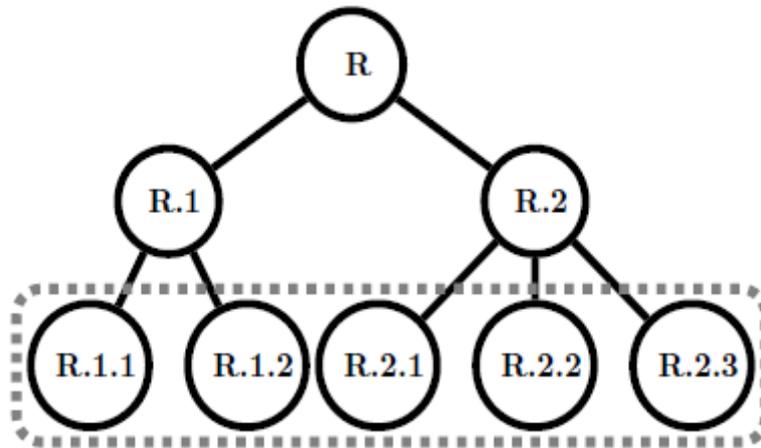


Figura 2.8: Transformação do problema hierárquico em problema de classificação plana

Fonte: Dias, 2015

Dias (2015) cita algumas desvantagens dessa abordagem:

- Uma delas está relacionada com o fato de o classificador perder a oportunidade de explorar os relacionamentos entre as classes presentes na hierarquia.
- Essa abordagem também é incapaz de lidar com problemas onde se deseja realizar previsões em qualquer nível da hierarquia, uma vez que somente classes dos nós folhas podem ser atribuídas às instâncias.

2.4.3.2. Predição Hierárquica com Algoritmos de Classificação Plana

Em várias bibliografias, esta abordagem é conhecida como abordagem local. Assim como a abordagem anterior, esta também transforma o problema de classificação hierárquica em problemas de classificação plana. Entretanto, nesta abordagem a estrutura hierárquica pode ser considerada em alguns níveis, pois cada nível da hierarquia é tratado como um problema de classificação independente. Assim, para cada nível, algoritmos de classificação plana podem ser então utilizados, como podemos ver na Figura 2.9. Ou seja, nessa abordagem, um ou mais classificadores independentes são construídos em cada nível da hierarquia de classes utilizando apenas informação local da taxonomia e do conjunto de exemplos de treinamento, como o conjunto de exemplos associados à classe

que o classificador representa. Além disso, cada classificador local “não tem conhecimento” da existência dos outros classificadores (METZ, 2011).

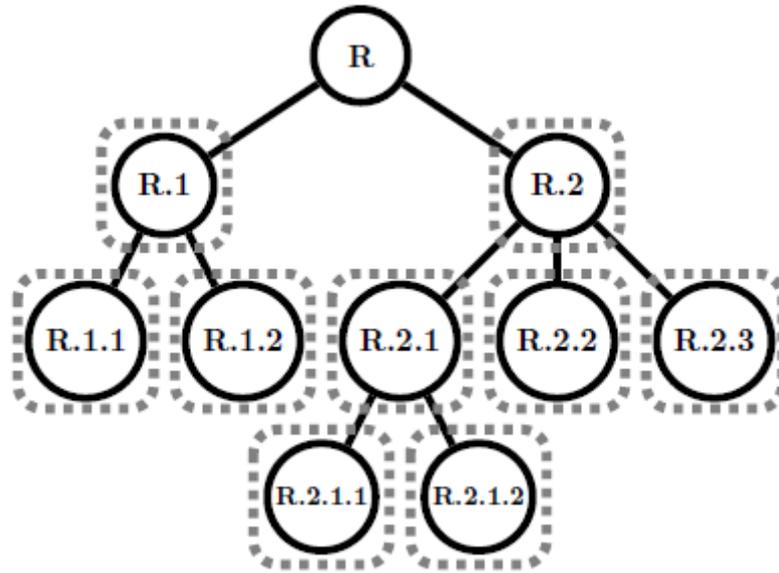


Figura 2.9: Classificação local
Fonte: Dias, 2015

2.4.3.3. Classificação Hierárquica *Top-Down*

A abordagem *top-down* consiste na construção de uma árvore de classificadores, através do treinamento de um ou mais classificadores para cada nível da hierarquia. Nesta abordagem, no nó raiz da hierarquia é treinado um classificador com todas as instâncias de treinamento. No nível seguinte, para cada classe é treinado um classificador, utilizando apenas um subconjunto de instâncias pertencentes às classes preditas pelo classificador, ou seja, o processo se repete recursivamente inspecionando apenas os classificadores referentes às classes imediatamente abaixo da(s) classe(s) predita(s) no passo anterior. Esse processo recursivo termina quando uma classe folha é encontrada (METZ, 2011). Na Figura 2.10 o passo a passo desse processo.

É importante salientar que na abordagem *top-down*, a geração de um classificador em um determinado nó da hierarquia só ocorre em nós nos quais haja a necessidade de um processo de decisão, ou seja, em nós que representam classes que possuam mais de uma subclasse. Nestes casos, há a necessidade de um processo de decisão (classificação) entre as subclasses para dar seguimento ao percurso *top-down* na hierarquia. Assim, configura-se a necessidade da indução de um classificador. Por outro

lado, nós internos que representam classes que possuem apenas uma subclasse e nós-folha não apresentam a necessidade de geração de classificadores (COSTA, 2008).

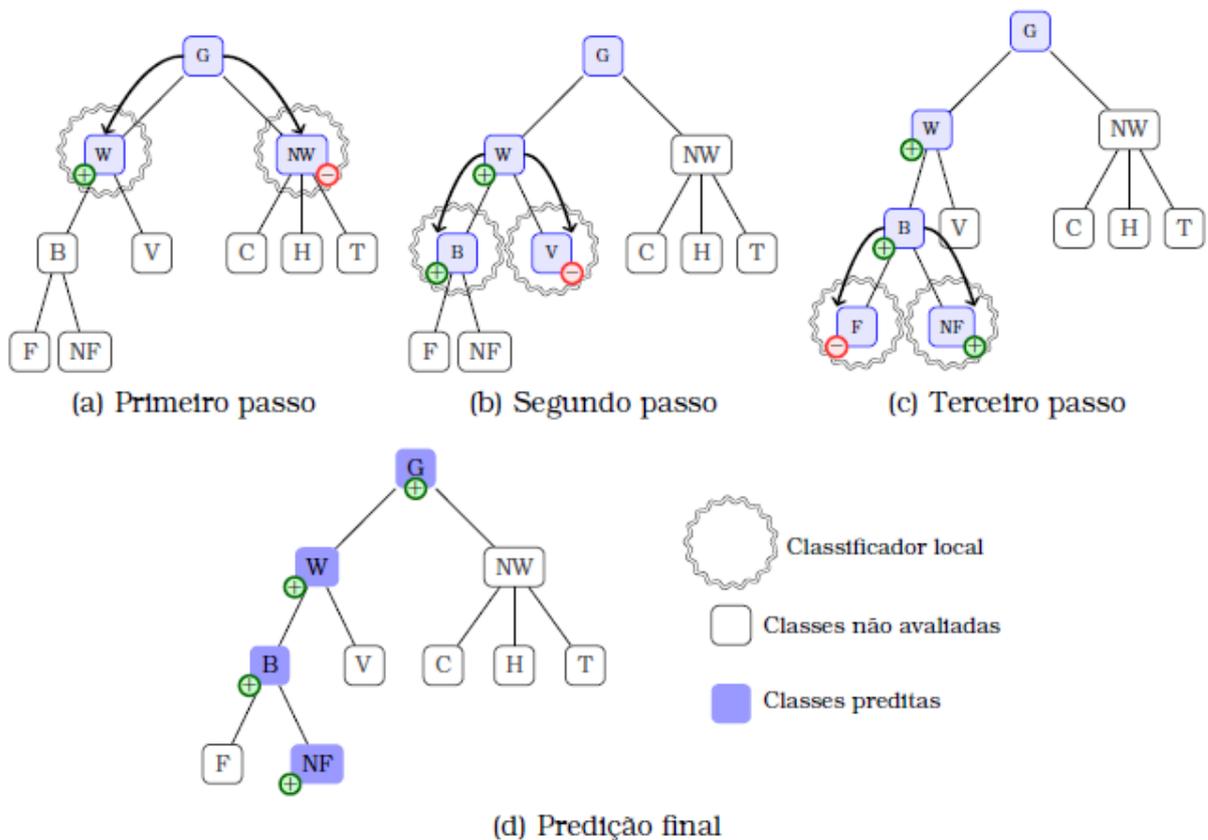


Figura 2.10: Abordagem *top-down*

Fonte: Metz, 2011

2.4.3.4. Classificação Hierárquica *Big-Bang*

Na abordagem *big-bang*, um único modelo de classificação é criado a partir de um conjunto de dados de treinamento em uma única rodada do algoritmo. Desse modo, uma vez que a hierarquia de classes como um todo é considerada, como podemos ver na Figura 2.11, o modelo gerado apresenta uma complexidade algorítmica mais alta que as abordagens anteriores. Após o treinamento do modelo de classificação, a predição da classe de uma nova instância é realizada em um único passo. Por esta razão, ao contrário das outras abordagens, a abordagem *big-bang* não pode usar técnicas de classificação plana em sua forma padrão, deve ser feita uma alteração em que a hierarquia fosse considerada como um todo.

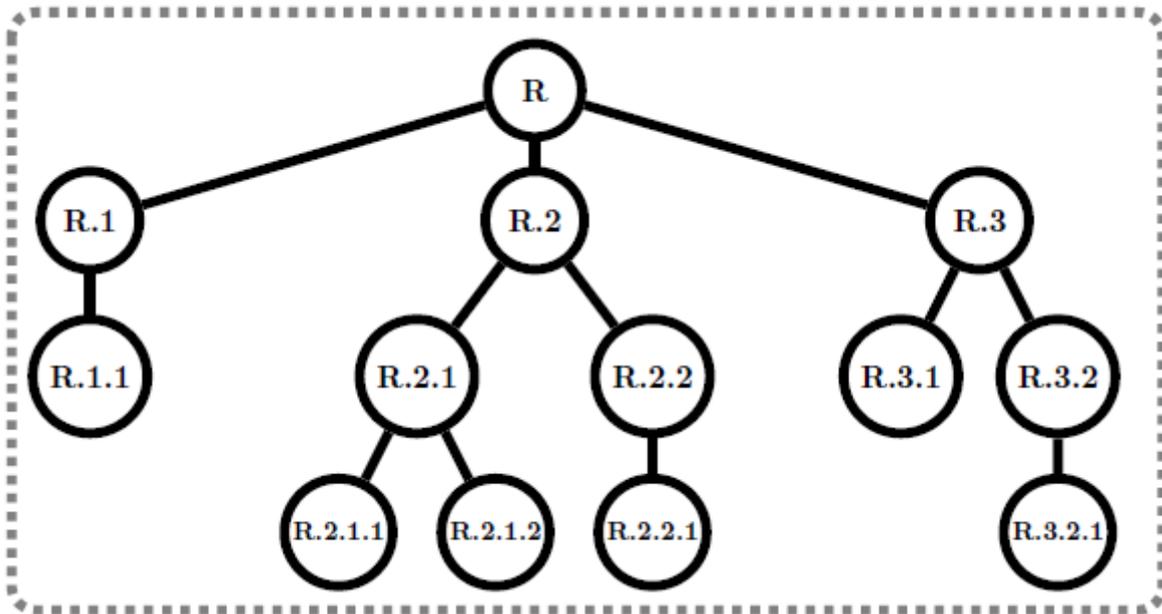


Figura 2.11: Abordagem *Big-Bang* (Global)

Fonte: Dias, 2015

Vale observar que a abordagem global perde a natureza modular das abordagens locais, ou seja, a característica de dividir a fase de treinamento em diversos processos, cada um considerando parte da hierarquia de classes. Portanto, o único classificador construído na abordagem global tende a ser mais complexo do que cada classificador individual construído nas abordagens locais (DIAS, 2015).

2.5. CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os conceitos de aprendizados de máquina, partindo dos conceitos básicos, seguido dos tipos de aprendizado, focando no aprendizado semissupervisionado, o qual foi mostrado as abordagens de descrição simples e multidescrição.

Após os tipos de aprendizado, foi conceituado as abordagens de classificação de dados com a classificação tradicional, multirrótulo e a hierárquica, apresentando as suas respectivas técnicas utilizadas neste trabalho.

Capítulo 3

TRABALHOS RELACIONADOS

Como já foi dito anteriormente, a área de aprendizagem de máquina está em constante evolução, resultando, assim, diversas pesquisas na mesma. Neste capítulo será apresentado trabalhos nesta área, focando no aprendizado semissupervisionado, classificação hierárquica multirrótulo e que utiliza o aprendizado semissupervisionado em problemas de classificação hierárquica multirrótulo.

3.1. APRENDIZADO SEMISSUPERVISIONADO

O aprendizado semissupervisionado tem como objetivo a rotulação de atributos não rotulados no conjunto de dados, a realização desta tarefa se dá através da experiência passada pelos dados rotulados deste mesmo conjunto, tal processo já foi relatado anteriormente neste trabalho, tanto como as duas vertentes desta metodologia de aprendizado, a visão simples e multivisão. Como nosso foco é o aprendizado multivisão, apenas os trabalhos que abordam este tema serão apresentados nesta sessão.

Em (BLUM; MITCHELL, 1998), foi proposto o primeiro algoritmo multivisão, o *Co-Training*, neste trabalho os autores dizem que cada visão é capaz de formar um bom classificador. Para a validação do método proposto, os experimentos foram realizados com páginas de Internet, mais especificamente, páginas dos departamentos de Ciência da Computação de quatro universidades, onde a primeira visão geraria um classificador com os atributos contidos nos textos das páginas e a segunda visão geraria o classificador a partir dos atributos contidos nos links de páginas que apontavam para as páginas dos departamentos. Os resultados obtidos utilizando o *Co-Training* foram comparados com

os resultados obtidos utilizando um método de aprendizado supervisionado. Onde, dos 16 exemplos rotulados e 800 não rotulados apresentados ao método semissupervisionado, o classificador gerado obteve 95% de taxa de precisão e o do aprendizado supervisionado, gerado apenas pelos 16 exemplos rotulados, obteve uma taxa de acerto de 88,9%.

Depois da criação do *Co-Training*, vários trabalhos surgiram propondo variações e melhorias no algoritmo proposto por (BLUM; MITCHELL, 1998), que por sua vez utiliza o *Naive Bayes* como algoritmo base. Muitas dessas variações modificam apenas o classificador base. Em (KIRITCHENKO; MATWIN, 2001), foi utilizado como classificador base o *Support Vector Machines*. Já em (ZHOU; LI, 2007), foi proposto um *Co-Training* que utiliza como algoritmo base o *k-Nearest Neighbor* (kNN), este chamado de COREG.

No algoritmo original de (BLUM; MITCHELL, 1998) existe um problema, denominado de problema dos pontos de contenção, que pode surgir quando os classificadores, em cada descrição, discordam com alta confiança na classificação. Em (BRAGA, 2010) algumas soluções são propostas, e uma delas é o algoritmo CoAL, este agregando o aprendizado ativo ao algoritmo *Co-Training*. Na fase experimental, foram utilizadas 11 bases de dados para a classificação de textos. Os resultados da avaliação experimental mostram que o CoAL consegue melhor desempenho que o *Co-Training* e que o algoritmo visão simples *Self-Training*.

Com o passar dos anos, novos métodos semissupervisionados surgem, em (XU; HE; MAN, 2012), foi apresentado um método que a diferença da Probabilidade Estimada da Classe (PEC) é utilizada para aumentar a diversidade dos classificadores, este foi implementado no *Co-Training*, denominado, assim, de *DCPE-Co-Training*. Este foi avaliado utilizando 13 bases de dados do *UCI Machine Learning Repository* e os classificadores bases foram: Naive Bayes, Rede Neural, KNN e Árvore de Decisão. O *DCPE-Co-Training* foi comparado com as técnicas de aprendizado supervisionado e semissupervisionado *Co-Training*, *Self-Training* e *Tri-Training*, sendo o método proposto pelo autor o que obteve os melhores resultados.

Em (CALDAS, 2017), foi proposto dois métodos semissupervisionados baseados na máquina de aprendizagem mínima utilizando o *Co-Training*, estes chamados de *Co-MLM* e *Fast Co-MLM*, para a fase de experimentos, foi utilizada uma base de dados

que contém um conjunto de dados de ações aeróbicas, chamada *UCF-DataPhone*, a mesma contém duas visões dos mesmos dados. Ao comparar as duas técnicas propostas pelo autor, ambos obtiveram resultados similares, porém, o *Fast Co-MLM* utilizou menos tempo de execução. Por fim, foram realizadas comparações entre o *Fast Co-MLM* e algoritmos mais conhecidos, como: *Co-Training*, *Co-Forest* e *TriTraining*. O método do autor apresentou resultados competitivos e o mesmo salienta que seu método pode ser estudado e que pode ser aprimorado.

3.2. CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO

Vimos no capítulo anterior que os dados de um conjunto de dados podem estar hierarquicamente dispostos e que esta relação pode ser considerada. Além disso, um único exemplo deste conjunto de exemplos pode pertencer a duas ou mais classes, ou seja, ser multirrótulo. Os problemas que abordam esses dois métodos, são conhecidos como problemas de classificação hierárquica multirrótulo. Este sendo tema de diversos trabalhos, alguns destes sendo abordados logo a seguir.

Em (TSOUMAKAS; KATAKIS; VLAHAVAS, 2008) é apresentado um algoritmo capaz de construir uma hierarquia de classes multirrótulo. O algoritmo chamado de HOMER (do inglês, *Hierarchy Of Multilabel classifiERs*) segue o paradigma “dividir para conquistar”, tendo como principal ideia a transformação de uma tarefa de classificação multirrótulo com um grande número de rótulos L em uma hierarquia em forma de árvore de tarefas de classificação multirrótulo, cada um lidando com um pequeno número $k \ll |L|$ de rótulos. O HOMER foi comparado com o método mais popular, o BR, os resultados mostraram que o HOMER provê mais acurácia de predição.

O trabalho de (CERRI; BARROS; CARVALHO, 2011) propõe um método local para classificação hierárquica multirrótulo utilizando redes neurais artificiais. O método é chamado HMC-LMLP (do inglês, *Hierarchical Multi-Label Classification with Local Multi-Layer Perceptron*), e utiliza uma rede *Multi-Layer Perceptron* (MLP) associada a cada nível da hierarquia. As predições feitas em um nível são então utilizadas como entrada para outra MLP responsável pelas predições no próximo nível. São utilizados dois algoritmos para o treinamento das MLPs, o algoritmo *Backpropagation* e o algoritmo *Resilient back-propagation*. O método é comparado com outros dois métodos. O Clus-HSC, método que explora os relacionamentos hierárquicos para construção de árvores de

decisão para cada nó hierárquico e o Clus-SC, método que gera árvores de decisão binárias para cada classe da hierarquia. Tais métodos são baseados no conceito de *Predictive Clustering Trees* (PCT). De acordo com os resultados experimentais, o método proposto obteve resultados preditivos competitivos, o que sugere as redes neurais artificiais como alternativas promissoras para tratar problemas de classificação hierárquica multirrótulo.

É comum encontrar trabalhos onde novos métodos são propostos e testados em bases de dados de predição de funções de gene, em (FENG; FU; ZHENG, 2017) foi proposto um novo método para HMC que possui quatro passos. Primeiramente é feita a preparação do conjunto de dados. O segundo passo é realizado um rebalanceamento dos dados, neste passo é utilizada uma abordagem chamada SMOTE, a qual é responsável por aliviar o problema de desbalanceamento dos dados. Depois é realizado o treinamento. E o quarto passo é realizada a predição nas instâncias de teste. O método proposto obteve resultados superiores em relação aos outros dois algoritmos utilizados, o TPR e o CLUS-HMC.

Um outro trabalho que utiliza bases de dados de predição de funções de gene é o de (ZHANG; SHAH; KAKADIARIS, 2017), neste, o autor propôs um novo local HMC *framework*, chamado de *Fully Associative Ensemble Learning*, o mesmo possui esse nome, pois, no seu modelo, a predição global de cada nó considera o relacionamento entre o nó atual e todos os outros nós. Este, construído para minimizar a perda entre a predição global de todos os exemplos de treinamento e suas correspondentes observações de rótulos verdadeiros. Além das bases de dados de predição de função de gene, também foram utilizadas bases com anotações de imagens e de reconhecimento visual. O *framework* proposto obteve resultados satisfatórios, porém, os autores, propõem uma melhoria de performance para trabalhos futuros.

3.3. APRENDIZADO SEMISSUPERVISIONADO EM PROBLEMAS DE CLASSIFICAÇÃO HIERÁRQUICA MULTIRRÓTULO

A utilização do aprendizado semissupervisionado vem mostrando ótimos resultados em pesquisas. A grande maioria delas abordam a classificação hierárquica e a classificação multirrótulo de forma separadas. Deixando, assim, a possibilidade que novos trabalhos tratem essas duas abordagens de forma conjunta.

Até o presente momento, poucos trabalhos utilizam o aprendizado semissupervisionado em problemas de classificação hierárquica multirrótulo. Em (SANTOS, 2012) são propostos algoritmos de aprendizado semissupervisionado multirrótulo e algoritmos de aprendizado semissupervisionado hierárquico multirrótulo, estes utilizam o algoritmo HMC (do inglês, *Hierarchical Multilabel Classification*) de (CERRI; CARVALHO, 2009). Os algoritmos que mais nos interessam no trabalho de (SANTOS, 2012) são: HMC-SSBR (do inglês, *Hierarchical Multilabel Classification using Semi-Supervised Binary Relevance*), HMC-SSLP (do inglês, *Hierarchical Multilabel Classification using Semi-Supervised Label Powerset*), HMC-SSRAkEL (do inglês, *Hierarchical Multilabel Classification using Semi-Supervised Random k-Labelsets*) e o HMC-SSRAkELd (do inglês, *Hierarchical Multilabel Classification using Semi-Supervised Random k disjoint Labelsets*). Com exceção deste último algoritmo citado, o HMC-SSRAkELd, no trabalho de (SANTOS; CANUTO, 2014) foi utilizado os demais métodos de aprendizado semissupervisionado visão simples em problemas de classificação hierárquica multirrótulo, neste caso, as bases de dados exploradas foram conjuntos de dados relacionados à funções de genes de fungos utilizados na fermentação de açúcar para a produção de etanol e bebidas alcoólicas. Ao comparar os métodos semissupervisionados visão simples com os métodos supervisionados (HMC-BR, HMC-LP e HMC-RAkEL) em problemas de classificação hierárquica multirrótulo, os resultados apontam uma performance similar, tal fato se comprova pela a não superioridade estatística na maioria dos casos, mostrando, assim, que o aprendizado semissupervisionado pode ser uma abordagem à ser utilizada quando se tem conjuntos de dados com poucos dados rotulados e muitos dados não rotulados.

3.4. CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados diversos trabalhos referentes às abordagens semissupervisionada multivisão, classificação hierárquica multirrótulo e semissupervisionada em problemas de classificação hierárquica multirrótulo, sendo, esta última, uma abordagem que contém poucos trabalhos disponíveis, abrindo, assim, o campo para diversas novas pesquisas. Tal escassez de contribuições só aumenta a importância deste trabalho de dissertação.

Capítulo 4

MATERIAIS E MÉTODOS

4.1. BASES DE DADOS

Para a fase experimental deste trabalho, foram utilizadas sete bases de dados biológicos, as quais estão dispostas em níveis hierárquicos e possui a característica multirrótulo. A subseção seguinte apresenta uma breve descrição dessas bases de dados. A Tabela 4.1 apresenta algumas características das bases de dados hierárquica multirrótulo.

4.1.1. Bases de Dados Hierárquica Multirrótulo

As bases de dados hierárquicas multirrótulo utilizadas neste trabalho estão relacionadas com as funções de genes do organismo *Saccharomyces cerevisiae* ou *Yeast* (levedura). Este fungo é frequentemente utilizado na fermentação de açúcar para a produção de etanol, e também na fermentação de trigo e cevada para a produção de bebidas alcoólicas. É um organismo modelo utilizado em muitos estudos. As funções dos genes deste organismo são organizadas em uma hierarquia estruturada como uma árvore de acordo com o esquema Funcat (<http://mips.gsf.de/projects/funcat>) de classificação desenvolvido pela MIPS (*Munich Information Center for Protein Sequences*) (MEWES *et al.*, 2000). Esses conjuntos de dados serviram como base para o trabalho de (CLARE, 2003). Neste trabalho foram utilizadas as bases de acordo com o trabalho de (CERRI, 2010), no qual, o autor realizou um pré-processamento para solucionar os problemas de

custo computacional e de memória, ambos detectados ao utilizar os conjuntos de dados originais.

Para a diminuição do número de classes dos conjuntos de dados, foram escolhidas, aleatoriamente, quatro classes do primeiro nível da hierarquia de classes, e todos os conjuntos foram limitados a possuírem no máximo quatro níveis hierárquicos. Assim, foram escolhidos, no primeiro nível, as classes “01”, “02”, “10” e “11” do esquema FunCat (CERRI, 2010). Ademais, o autor realizou procedimentos para garantir que, para cada classe, houvesse exemplos suficientes para a fase de treinamento e o mesmo utilizou a estratégia de imputação de valores ausentes para que o conjunto de dados não tivesse valores faltosos. Contudo, neste trabalho, iremos abordar até o terceiro nível da hierarquia.

Tabela 4.1: Resumo das Bases de Dados Hierárquica Multirrótulo

Base	N. Atributos	Rótulos (1N)	Rótulos (2N)	Rótulos (3N)
Cellcycle	77	4	22	70
Church	27	4	22	70
Derisi	63	4	22	70
Eisen	79	4	22	66
Gasch1	173	4	22	70
Gasch2	52	4	22	70
SPO	80	4	22	70

4.1.1.1. *Cellcycle*

Esta base de dados foi coletada de um *microarray* de dados obtidos a partir de 4 séries de experimentos (SPELLMAN *et al.*, 1998). O conjunto de dados utilizado possui 77 atributos reais.

4.1.1.2. *Church*

Esta base de dados foi coletada de um *microarray* de dados (ROTH *et al.*, 1998). O conjunto de dados utilizado possui 27 atributos reais.

4.1.1.3. *Derisi*

Esta base de dados foi coletada de um *microarray* de dados (DERISI; IYER; BROWN, 1997). O conjunto de dados utilizado possui 63 atributos reais.

4.1.1.4. *Eisen*

Esta base de dados foi coletada de um *microarray* de dados obtidos a partir de 4 séries de experimentos (EISEN *et al.*, 1998). O conjunto de dados utilizado possui 79 atributos reais.

4.1.1.5. *Gasch1*

Esta base de dados foi coletada de um *microarray* constituído de 173 atributos reais, obtidos de (GASCH *et al.*, 2000).

4.1.1.6. *Gasch2*

Esta base de dados foi coletada de um *microarray* constituído de 52 atributos de valores reais, obtidos de (GASCH *et al.*, 2001).

4.1.1.7. SPO

Esta base de dados foi coletada de um *microarray* constituído de 80 atributos, que em sua maioria são valores reais, obtidos do trabalho de (CHU *et al.*, 1998).

4.2. MÉTRICAS DE AVALIAÇÃO

A avaliação de classificadores multirrótulo requer medidas de avaliação distintas das utilizadas no caso de problemas de classificação unirrótulo. Ao contrário dos problemas unirrótulo, em que a classificação de um exemplo é correta ou incorreta, em um problema multirrótulo a classificação de um exemplo pode ser parcialmente correta ou incorreta. Isto pode ocorrer quando um classificador atribui corretamente um exemplo a ao menos uma das classes a que ele este pertence, mas não atribui a todas as classes a

que o exemplo pertence. Além disso, o classificador pode também atribuir um exemplo a uma ou mais classes a que ele não pertence (CERRI; CARVALHO, 2009).

Várias medidas têm sido propostas na literatura para a avaliação de classificadores multirrótulo. De acordo com (TSOUMAKAS; KATAKIS; VLAHAVAS, 2009), estas medidas podem ser categorizadas em dois grupos: medidas baseadas em bipartição e medidas baseadas em ranking. Algumas destas medidas baseadas em bipartição avaliam as bipartições sobre todos os exemplos da base de dados de avaliação. Outras medidas baseadas em bipartição decompõem o processo de avaliação em avaliações separadas para cada rótulo. Além disso, as medidas baseadas em ranking avaliam os rankings com respeito à precisão da base de dados multirrótulo.

Nas subseções seguintes, serão apresentadas algumas medidas de avaliação que serão utilizadas neste trabalho.

4.2.1. Métricas Baseadas em Bipartição

Foram utilizadas, nesse trabalho, as seguintes medidas baseadas em bipartição: *Hamming Loss* (HLoss), *Subset Accuracy* (SAcc) e *Accuracy* (Acc).

4.2.1.1. *Hamming Loss*

O *Hamming Loss* avalia a porcentagem correta de rótulos não preditos e rótulos preditos incorretamente. Dessa maneira, esta medida considera a frequência com que um par exemplo/rótulo é classificado de maneira incorreta, ou seja, cada exemplo é associado a um rótulo errado ou um rótulo não predito. Esta medida define que o melhor desempenho é alcançado quando o resultado é igual a zero, isto é, quanto mais o valor se aproxima a zero, melhor será seu desempenho (RODRIGUES, 2014). A Figura 4.1 mostra a formula matemática da métrica *Hamming Loss*.

$$HLoss = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \Delta Z_i|}{M}$$

Figura 4.1: Equação da fórmula matemática da métrica de *Hamming Loss*

Fonte: Rodrigues, 2014

4.2.1.2. *Subset Accuracy*

Esta métrica é uma medida de acurácia muito restrita, considerando como correto se todos os rótulos preditos por um classificador são corretos. A Figura 4.2 mostra a formula matemática da métrica *Subset Accuracy*.

$$SAcc = \frac{1}{N} \sum_{i=1}^N I(|Z_i| - |Y_i|)$$

Figura 4.2: Equação da fórmula matemática da métrica de *Subset Accuracy*

Fonte: Santos, 2012

4.2.1.3. *Accuracy*

A métrica de *Accuracy* avalia a porcentagem de verdadeiros positivos fora do total de verdadeiros positivos, falsos positivos e negativos. A Figura 4.3 mostra a formula matemática da métrica *Accuracy*.

$$Acc = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap Z_i|}{|Y_i \cup Z_i|}$$

Figura 4.4: Equação da fórmula matemática da métrica de *Accuracy*

Fonte: Santos, 2012

4.2.2. Métricas Baseadas em Ranking

Utilizou-se, nesse trabalho, as seguintes medidas baseadas em *ranking*: *One-Error* (1-Err), *Average Precision* (AvPrec) e *Ranking Loss* (RLoss).

4.2.2.1. *One-Error*

Esta métrica avalia a frequência de rótulos no topo do *ranking* que não estavam no conjunto de rótulos corretos. O melhor desempenho é alcançado quando esta medida é igual a 0. Assim, quanto menor o valor alcançado, melhor é o desempenho. A Figura 4.4 mostra a formula matemática da métrica *One-Error*.

$$1Error = \frac{1}{N} \sum_{i=1}^N I(\operatorname{argmin}_{y \in L} r_i(y))$$

Figura 4.4: Equação da fórmula matemática da métrica de *One-Error*

Fonte: Cherman, 2013

4.2.2.2. *Average Precision*

Esta métrica representa a precisão média tomada para todos os possíveis rótulos e pode avaliar algoritmos como um todo. A *Average Precision* mede a fração média de rótulos ranqueado acima de um rótulo particular $l \in Y_i$ que está atualmente em Y_i . O melhor desempenho é alcançado quando a precisão média é igual a 1. Assim, quanto maior o valor da precisão média, melhor é o desempenho. A Figura 4.5 mostra a fórmula matemática da métrica *Average Precision*.

$$AveragePrecision = \frac{1}{N} \sum_{i=1}^N \frac{1}{|Y_i|} \sum_{y \in Y_i} \frac{|(y' \in Y_i : r_i(y') \leq r_i(y))|}{r_i(y)}$$

Figura 4.5: Equação da fórmula matemática da métrica de *Average Precision*

Fonte: Cherman, 2013

4.2.2.3. *Ranking Loss*

Esta métrica expressa quantas vezes rótulos irrelevantes são ranqueados acima de rótulos relevantes, ou seja, quantas vezes um rótulo $y' \in Y_i$ tem um ranking pior que um rótulo $y' \in L - Y_i$. A Figura 4.6 mostra a fórmula matemática da métrica *Ranking Loss*.

$$Ranking-Loss = \frac{1}{N} \sum_{i=1}^N \frac{1}{|Y_i| | \bar{Y}_i |} |(y_a, y_b) : r_i(y_a) < r_i(y_b), (y_a, y_b) \in Y_i \times (L - Y_i)|$$

Figura 4.6: Equação da fórmula matemática da métrica de *Ranking Loss*

Fonte: Cherman, 2013

4.2.3. Hierarchical Loss

Para avaliação da classificação hierárquica multirrótulo ainda existem poucas métricas disponíveis, o que abre campo para várias pesquisas exclusivas neste sentido. Porém, neste trabalho, utilizamos a *Hierarchical Loss* (HiLoss), proposta no trabalho de (CESA-BIANCHI; GENTILE; ZANIBONI, 2006), esta baseia-se na noção de ancestralidade e descendência das classes preditas no momento dos cálculos da avaliação, sendo utilizada a noção intuitiva de que sempre que um erro de classificação é cometido em um nó da hierarquia de classes, não deve haver penalizações adicionais para erros cometidos na subárvore desse nó. Desse modo, se um erro é cometido em um nó i , erros adicionais cometidos na subárvore rotulada por i não são importantes. A Figura 4.7 mostra a formula matemática da métrica *Hierarchical Loss*.

$$\text{H-Loss} = \frac{1}{m} \sum_{i=1}^m |\{\lambda : \lambda \in Y_i \Delta Z_i, \text{anc}(\lambda) \cap (Y_i \Delta Z_i) = \emptyset\}|$$

Figura 4.7: Equação da fórmula matemática da métrica de *Hierarchical Loss*

Fonte: Tsoumakas; Katakis; Vlahavas, 2010

4.3. CONFIGURAÇÃO DOS EXPERIMENTOS

Para cada um dos métodos de classificação hierárquica multirrótulo supervisionado e semissupervisionado foram aplicados 4 diferentes algoritmos como classificadores base: k Vizinhos Mais Próximos (k-NN), Algoritmo de Indução de Árvores de Decisão (J48), *Naive Bayes* (NB) e *Extended Repeated Incremental Pruning* (JRip). Estes classificadores foram escolhidos por serem bastante utilizados na literatura. Todos os resultados apresentados neste trabalho referem-se as médias dos melhores resultados obtidos entre os 4 classificadores base.

Os experimentos utilizando os métodos de classificação semissupervisionada necessitam da definição de dois parâmetros adicionais: o percentual inicial de exemplos não rotulados e o percentual de exemplos incorporados a cada iteração. Com relação ao percentual inicial de exemplos não rotulados foram utilizados dois valores: 50 e 75% dos exemplos da base utilizada, pois objetiva-se investigar o comportamento dos resultados com uma baixa dimensionalidade dos dados comparado com uma dimensionalidade

maior. Por sua vez, com relação ao percentual de exemplos incorporados a cada iteração foram utilizados quatro valores: 20, 30, 40 e 50%. A Figura 4.8 mostra a árvore da configuração dos experimentos semissupervisionados.

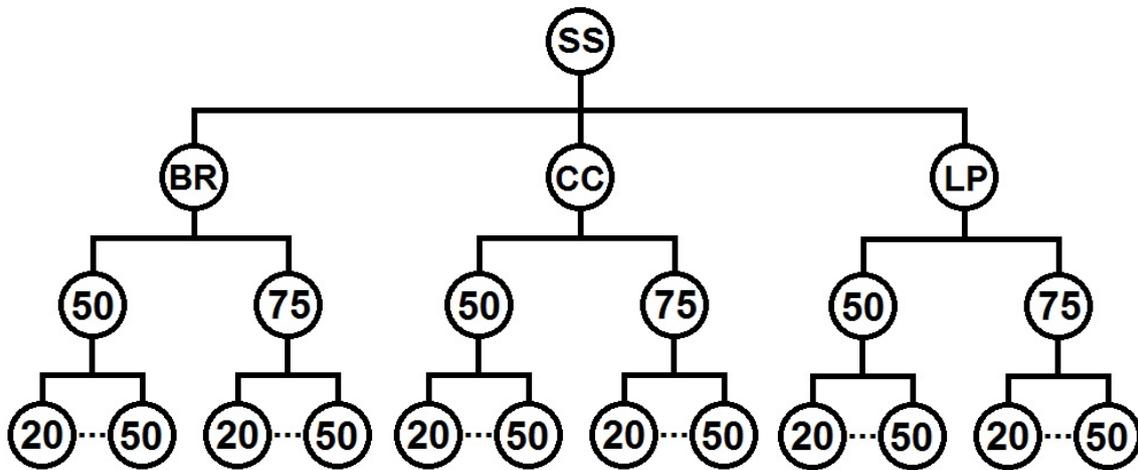


Figura 4.8: Árvore da Configuração dos Experimentos Semissupervisionados

Todos os experimentos foram conduzidos usando a metodologia de validação cruzada com 10 grupos, $K=10$ (*10-fold Cross-Validation*). Vale ressaltar, que em relação as configurações internas dos classificadores durante os experimentos, foram utilizados o padrão *default* definidos pelo *Waikato Environment for Knowledge Analysis* (WEKA), que é uma ferramenta que possui uma coleção de algoritmos de aprendizagem de máquina para tarefas de mineração de dados (WEKA, 2017). Os resultados a seguir, são os melhores resultados obtidos.

Para a análise dos resultados obtidos pelos diferentes métodos de aprendizado foram utilizados dois testes bastante conhecidos: o Teste de Wilcoxon e o Teste de Friedman. O primeiro foi utilizado para as comparações pareadas, ou seja, entre duas amostras, enquanto o segundo foi utilizado nas comparações entre três ou mais amostras. Foi utilizado o nível de confiança de 95% ($\alpha = 0,05$) durante todos os testes estatísticos.

4.4. TESTES ESTATÍSTICOS

Testes estatísticos são utilizados para analisar a significância estatística das diferenças apresentadas pelos resultados experimentais dos métodos de classificação utilizados neste trabalho. Os testes estatísticos fornecem um respaldo científico às

pesquisas para que estas tenham validade e credibilidade. Em estatística, diz-se que um resultado é significativo se for improvável que tenha ocorrido por acaso. Dessa forma, busca-se com a utilização de um teste estatístico, determinar se a diferença encontrada entre os sistemas indica superioridade de um sistema em relação ao outro. Para isso formulam-se duas hipóteses, a primeira, hipótese nula, é tida como verdadeira até que as provas estatísticas indiquem o contrário. A segunda, hipótese alternativa, deve ser oposta à primeira. Dessa forma, as hipóteses sendo contraditórias não poderão ser simultaneamente verdadeiras. Para o problema proposto neste trabalho, as hipóteses levantadas são:

- Hipótese nula (H_0): Não existe diferença estatística entre o desempenho dos sistemas, ou seja, não pode-se dizer que um sistema seja estatisticamente superior ao outro;
- Hipótese alternativa (H_1): Os sistemas possuem desempenhos estatisticamente diferentes, podendo-se dizer que um sistema é estatisticamente superior ao outro.

Deve-se levar em consideração que para a execução do teste estatístico deve-se definir previamente o nível de significância ou nível de confiança, que representa a probabilidade de que o resultado estatístico encontrado não seja verdadeiro. Em geral, são aceitos como significativos os níveis $P = 0,05$ e $P = 0,01$, ou seja, no caso de rejeitar-se a hipótese nula, o faz com 95% ou 99% de confiança sobre sua correção. Neste trabalho foi adotado o nível de confiança de 95 % ($\alpha = 0,05$). Existe uma série de testes estatísticos que podem auxiliar as pesquisas. Neste trabalho, foram utilizados dois testes estatísticos bastante utilizados pela comunidade científica: o Teste de Wilcoxon e Teste de Friedman. O Teste de Wilcoxon é um método não-paramétrico para comparação de duas amostras pareadas. A princípio são calculados os valores numéricos da diferença entre cada par, sendo possível três condições: aumento (+), diminuição (-) ou igualdade (=). Uma vez calculadas todas as diferenças entre os valores obtidos para cada par de dados, essas diferenças são ordenadas pelo seu valor absoluto (sem considerar o sinal), substituindo-se então os valores originais pelo posto que ocupam na escala ordenada (WILCOXON, 1945). O teste da hipótese de igualdade entre os grupos é baseado na soma dos postos das diferenças negativas e positivas. Este teste para dados pareados, ao invés de considerar apenas o sinal das diferenças entre os pares, considera o valor dessas diferenças, sendo assim um teste não-paramétrico dos mais poderosos e populares. Já o Teste de Friedman é amplamente utilizado para comparar três ou mais amostras (conjunto

de resultados). O Teste de Friedman é um teste não-paramétrico que utiliza os rankings dos algoritmos em cada um dos conjunto de resultados. O algoritmos que obtiver o melhor desempenho recebe a posição 1 do ranking, o segundo melhor recebe a posição 2 e assim por diante (FRIEDMAN, 1937). Se ao utilizar o Teste de Friedman for constatado uma relevância estatística, em seguida é realizado o pós-teste com o Teste de Wilcoxon para que se possa identificar quais amostras possuem superioridade estatística.

4.5. CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentadas as bases de dados hierárquicas multirrótulo utilizadas neste trabalho, as métricas de avaliação, logo após foram explicadas as configurações dos experimentos e, por fim, foi-se falado sobre os testes estatísticos utilizados.

Capítulo 5

RESULTADOS DOS EXPERIMENTOS

Conforme citado anteriormente, neste trabalho é proposto a utilização de técnicas de aprendizado semissupervisionado em problemas de classificação hierárquica multirrótulo existentes. Com o objetivo de possibilitar a análise do impacto desta proposta no comportamento dos métodos de classificação hierárquica multirrótulo, experimentos foram realizados utilizando os métodos com a abordagem supervisionada e a semissupervisionada, aplicados em conjuntos de dados biológicos relativos à predição de funções de proteínas.

Neste capítulo, será realizada uma investigação comparativa do comportamento dos métodos de classificação hierárquica multirrótulo com abordagens de aprendizado supervisionado, semissupervisionado visão simples e semissupervisionado multidescrição. Além das diferentes abordagens de aprendizado e dos diferentes métodos de classificação hierárquica multirrótulo, os experimentos utilizaram diferentes valores para alguns parâmetros, tais como: percentual inicial de exemplos não rotulados e percentual de incorporação a cada iteração. Desse modo, em virtude do grande número de métodos e parâmetros variáveis utilizados, a análise experimental será realizada de uma perspectiva mais geral, de acordo com a abordagem de aprendizado utilizada no processo de construção dos classificadores.

Neste contexto, todos os resultados apresentam a média e, em alguns, também o desvio padrão dos vários métodos supervisionados e semissupervisionados utilizados de acordo com a perspectiva a ser analisada, tendo sido destacado os melhores resultados em cada uma das medidas. Além disso, testes estatísticos foram realizados para analisar a

relevância estatística das diferenças observadas nos resultados dos diferentes métodos e abordagens de aprendizado (ver Seção 4.4). Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **FRI** marcados em negrito. A coluna **WIL**, nas tabelas, exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas ou o valor de relevância do ponto de vista estatístico, no caso de comparações pareadas, nestes casos, os valores que apresentaram relevância também tiveram seus valores marcados em negrito. Vale ressaltar que, para algumas métricas de avaliação, os melhores resultados foram obtidos com valores próximos a zero, enquanto que, para outras métricas, os melhores resultados foram obtidos com valores próximos a um. Portanto, ao lado de cada métrica, foi adicionado o símbolo ↓ (o menor valor significa o melhor) e ↑ (o maior valor significa o melhor) para representar o comportamento das métricas de avaliação.

Nesta perspectiva, os resultados dos experimentos serão apresentados da seguinte forma para o nível 1 (Seção 5.1), nível 2 (Seção 5.2) e nível 3 (Seção 5.3):

1. Inicialmente, os resultados dos métodos que utilizam a abordagem de aprendizado supervisionado serão apresentados da seguinte forma:
 - a. Técnicas de classificação multirrótulo.
2. Depois, os resultados dos métodos que utilizam a abordagem de aprendizado semissupervisionado serão apresentados da seguinte forma para visão simples e para multidescrição, onde cada tópico a seguir é apresentado primeiramente na visão simples e, logo em seguida, o mesmo é mostrado para o multidescrição:
 - a. Técnicas de classificação multirrótulo;
 - b. Percentuais de dados não rotulados iniciais;
 - c. Percentuais de incorporação por iteração.
3. Após isso, serão apresentados os resultados dos métodos de classificação supervisionada e semissupervisionadas, de modo a possibilitar uma análise comparativa entre as abordagens.

Por fim (Seção 5.4), análises, de cada técnica de aprendizado, são apresentadas nos seus três níveis hierárquicos, objetivando apresentar seus desempenhos ao passar de um nível para o outro.

Desse modo, as Seções 5.1, 5.2 e 5.3 apresentam os resultados dos métodos de classificação hierárquica multirrótulo no primeiro, no segundo e no terceiro nível da hierarquia, respectivamente. A Seção 5.4 apresenta uma análise de desempenho entre níveis. Por fim, a Seção 5.5 discute, em linhas gerais, os resultados obtidos. As tabelas, contendo os resultados dos experimentos destas seções citadas, se encontram no Apêndice A.

5.1. MÉTODOS HIERÁRQUICOS MULTIRRÓTULO: PRIMEIRO NÍVEL

Nesta seção, são apresentados os resultados dos métodos de classificação hierárquica multirrótulo supervisionada e semissupervisionadas. É importante enfatizar que todos os resultados desta seção referem-se aos resultados obtidos no primeiro nível da hierarquia. Inicialmente, serão apresentados os resultados dos métodos de classificação hierárquica multirrótulo supervisionada sob a perspectiva das técnicas multirrótulo no contexto da classificação hierárquica. Após isso, serão apresentados os resultados dos métodos de classificação hierárquica multirrótulo semissupervisionadas sob a perspectiva das técnicas multirrótulo no contexto da classificação hierárquica, do percentual de dados rotulados iniciais e, após isso, da perspectiva do percentual de incorporação do iteração. O objetivo da análise sob várias perspectivas distintas reside no estudo do impacto da escolha dos valores dos parâmetros no comportamento dos classificadores. Ao realizar esta análise, espera-se ter um indicativo de quais valores dos parâmetros apresentam os melhores resultados, considerando o universo estudado. Por fim, será realizada uma análise comparativa dos resultados obtidos pelos métodos de classificação hierárquica multirrótulo supervisionada e dos métodos de classificação hierárquica multirrótulo semissupervisionadas.

5.1.1. Melhores Resultados dos Métodos BR, CC e LP no Contexto Supervisionado

Analisando a Tabela A.1, é possível observar que os experimentos utilizando a classificação hierárquica multirrótulo supervisionada com os métodos BR, CC e LP, o método LP apresentou melhor desempenho em 50,00% dos casos (21/42), seguido do CC com 28,57% dos casos (12/42) e o BR com 21,43% dos casos (9/42). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 45,24% dos casos (19/42). O LP mostrou-se superior do ponto de vista estatística em

21,43% dos casos (9/42), seguido pelo CC com 19,05% dos casos (8/42) e BR com 4,76% dos casos (2/42). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 39,47% dos casos (15/38) o Teste de Wilcoxon encontrou relevância estatística. A Figura 5.1 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo supervisionado no primeiro nível da hierarquia, ao analisar a mesma, vê-se que, neste caso, o método LP obteve o melhor desempenho entre os métodos utilizados, isso se deve ao fato do LP considerar as relações entre as classes e a quantidade de subconjuntos de classes gerados pelo mesmo tenha ajudado na melhoria da geração do classificador, porém, não pode-se verificar relevância estatística na maioria dos casos.

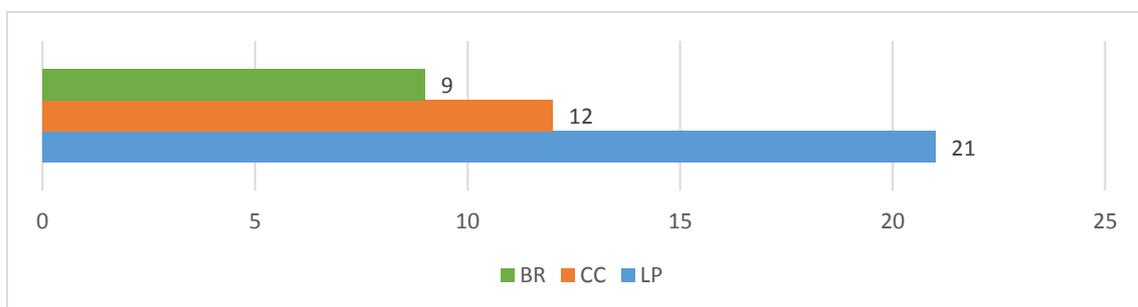


Figura 5.1: Métodos de Classificação Hierárquica Multirrótulo Supervisionados no Nível 1 da Hierarquia

5.1.2. Melhores Resultados dos Métodos BR, CC e LP no Contexto Semissupervisionado Visão Simples

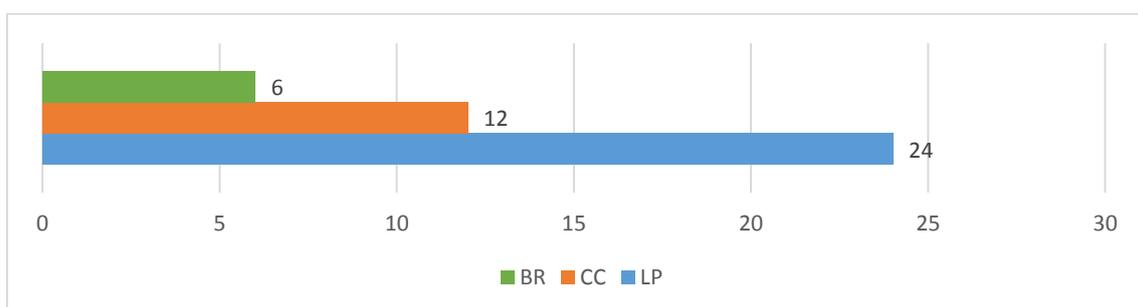


Figura 5.2: Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Visão Simples no Nível 1 da Hierarquia

A Figura 5.2 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo semissupervisionado visão simples no primeiro nível da hierarquia, resultados obtidos após uma investigação na Tabela A.2, da qual se tornou possível observar que os experimentos utilizando a classificação hierárquica multirrótulo semissupervisionada visão simples com os métodos BR, CC e LP, o método LP apresentou melhor desempenho em 57,14% dos casos (24/42), seguido do CC com 28,57% dos casos (12/42) e o BR com 14,29% dos casos (6/42). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 52,38% dos casos (22/42). O LP mostrou-se superior do ponto de vista estatística em 30,95% dos casos (13/42), seguido pelo CC com 19,05% dos casos (8/42) e BR com 2,38% dos casos (1/42). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 43,18% dos casos (19/44) o Teste de Wilcoxon encontrou relevância estatística. Ao analisar a Figura 5.2, vê-se que, neste caso, o método LP obteve o melhor desempenho entre os métodos utilizados, isso se deve ao fato do LP considerar as relações entre as classes e a quantidade de subconjuntos de classes gerados pelo mesmo tenha ajudado na melhoria da geração do classificador, contudo, pode-se verificar relevância estatística na maioria dos casos.

5.1.3. Melhores Resultados dos Métodos BR, CC e LP no Contexto Semissupervisionado Multivisão

Investigando a Tabela A.3, é possível observar que os experimentos utilizando a classificação hierárquica multirrótulo semissupervisionada multivisão com os métodos BR, CC e LP, o método LP apresentou melhor desempenho em 64,29% dos casos (27/42) seguido do BR com 21,43% dos casos (9/42) e o CC com 14,29% dos casos (6/42). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 45,24% dos casos (19/42). O LP mostrou-se superior do ponto de vista estatística em 38,10% dos casos (16/42), seguido pelo BR com 4,76% dos casos (2/42) e CC com 2,38% dos casos (1/42). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 63,16% dos casos (24/38) o Teste de Wilcoxon encontrou relevância estatística. A Figura 5.3 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo semissupervisionado multivisão no primeiro nível da hierarquia, ao analisar a mesma, vê-se que, neste caso, o método LP obteve o melhor desempenho entre os métodos utilizados, isso se deve ao fato do LP

considerar as relações entre as classes e a quantidade de subconjuntos de classes gerados pelo mesmo tenha ajudado na melhoria da geração do classificador, porém, não pode-se verificar relevância estatística na maioria dos casos.

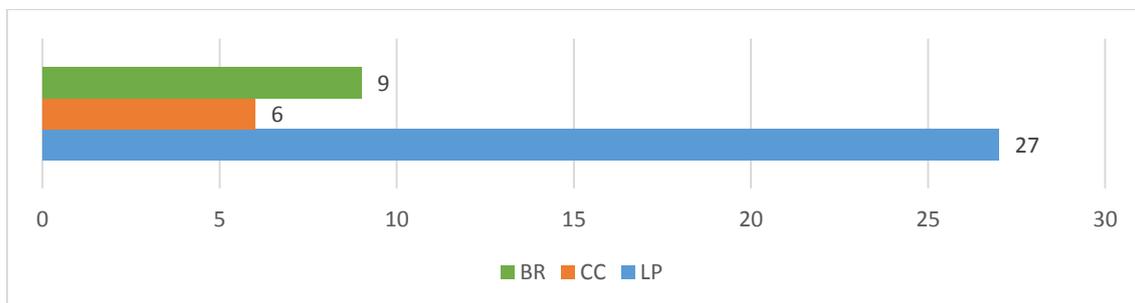


Figura 5.3: Métodos de Classificação Hierárquica Multirrotulo Semissupervisionado Multivisão no Nível 1 da Hierarquia

5.1.4. Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados

Analisando a Tabela A.4, é possível observar que o valor de 50% de dados não rotulados iniciais foi superior em 97,62% dos casos (41/42), enquanto que o 75% de não rotulados, ou seja, apenas 25% dos dados rotulados iniciais, obteve 2,38% dos casos (1/42). Ao realizar a os testes estatísticos, não foi verificada a relevância estatística em nenhum dos casos. Com isso, podemos afirmar que, neste caso, os valores utilizados não apresentam impacto relevante nos resultados apresentados, porém, o valor de 50% obteve melhor desempenho. A Figura 5.4 ilustra o desempenho do método semissupervisionado visão simples para percentual inicial de exemplos não rotulados no primeiro nível da hierarquia.

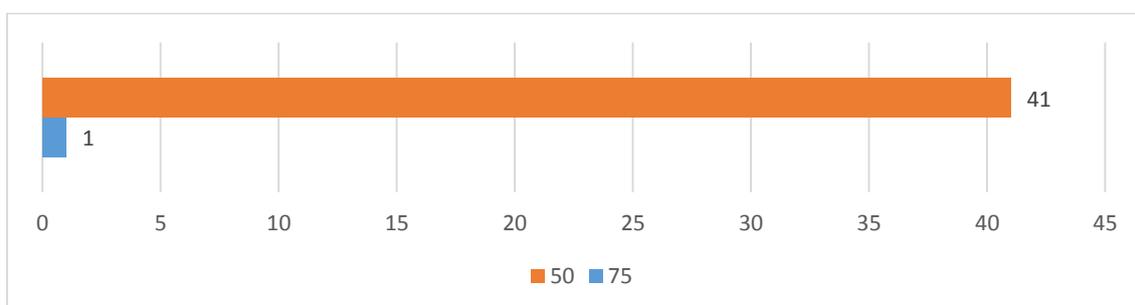


Figura 5.4: Método Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados no Nível 1 da Hierarquia

5.1.5. Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados

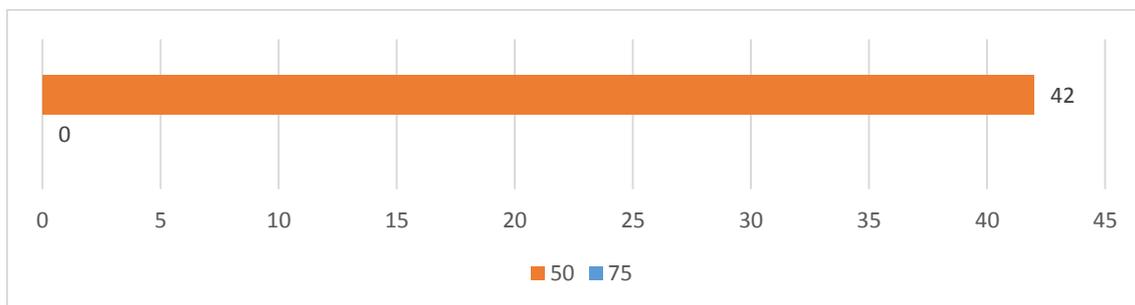


Figura 5.5: Método Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados no Nível 1 da Hierarquia

A Figura 5.5 ilustra o desempenho do método semissupervisionado multivisão para percentual inicial de exemplos não rotulados no primeiro nível da hierarquia, resultado obtido após verificar a Tabela A.5, da qual é possível observar que o valor de 50% de dados não rotulados iniciais foi superior em 100% dos casos (42/42), enquanto que o 75% de não rotulados não foi superior em nenhum caso. Ao aplicar o Teste de Wilcoxon, verificou-se relevância estatística em 21,43% dos casos (9/42). Com isso, pode-se afirmar que quanto maior o número de dados rotulados iniciais, melhor o desempenho dos classificadores, porém, não pode-se verificar relevância estatística na maioria dos casos.

5.1.6. Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados

Investigando a Tabela A.6, com os valores das porcentagens de incorporação por iteração, observa-se que em 28,57% dos casos (12/42) os valores de 30 e 50% ficaram empatados com os melhores resultados, seguidos do percentual de 40% com 26,19% dos casos (11/42) e o de 20% com 16,67% dos casos (7/42). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 2,38% dos casos (1/42), neste caso, foi o valor de 40% que obteve superioridade estatística em relação aos valores de 20 e 30%. A Figura 5.6 ilustra o desempenho do método semissupervisionado visão simples para percentual incorporação por iteração de novos

exemplos rotulados no primeiro nível da hierarquia. Ao observar a Figura 5.6, vê-se que quanto maior o percentual de iteração por ciclo, melhor o desempenho do classificador, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

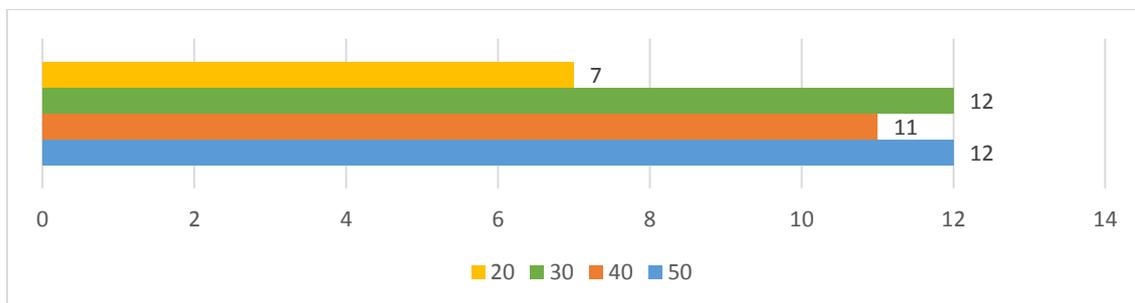


Figura 5.6: Método Semissupervisionado Visão Simples Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 1 da Hierarquia

5.1.7. Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados

Averiguando a Tabela A.7, com os valores das porcentagens de incorporação por iteração, observa-se que em 38,10% dos casos (16/42) o valor de 40% ficou com o melhor resultado, seguido do percentual de 50% com 33,33% dos casos (14/42), o 20% com 16,67% dos casos (7/42) e o de 30% com 11,90% dos casos (5/42). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 2,38% dos casos (1/42), neste caso, foi o valor de 20% que obteve superioridade estatística em relação ao valor de 50%. A Figura 5.7 ilustra o desempenho do método semissupervisionado multivisão para percentual incorporação por iteração de novos exemplos rotulados no primeiro nível da hierarquia. Ao observar a Figura 5.7, vê-se que quanto maior o percentual de iteração por ciclo, melhor o desempenho do classificador, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

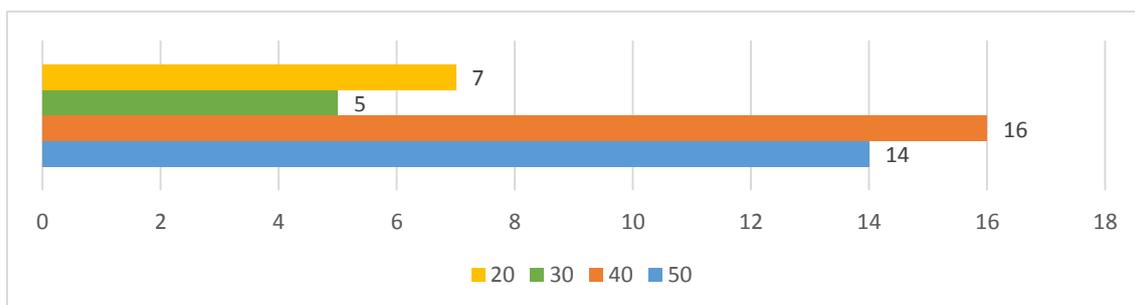


Figura 5.7: Método Semissupervisionado Multivisão Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 1 da Hierarquia

5.1.8. Abordagem Semissupervisionada Multivisão *versus* Semissupervisionada Visão Simples *versus* Supervisionada

Explorando a Tabela A.8, observa-se que o método supervisionado foi o melhor com 73,81% dos casos (31/42), seguido do semissupervisionado Co-Training com 19,05% dos casos (8/42) e o semissupervisionado Self-Training com 7,14% dos casos (3/42). Ao aplicar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 2,38% dos casos (1/42), neste caso, o semissupervisionado *Co-Training* obteve superioridade estatística em relação ao aprendizado supervisionado. A Figura 5.8 ilustra o desempenho dos métodos semissupervisionado multivisão *versus* semissupervisionado visão simples *versus* supervisionado no primeiro nível da hierarquia. Ao analisar a Figura 5.8, pode-se observar que o aprendizado supervisionado foi superior que o métodos semissupervisionados, tal superioridade é oriunda do fato de que o método de aprendizado supervisionado realizava a fase de treinamento com todos os exemplos disponíveis, gerando, assim, classificadores com melhor desempenho comparados com os dos métodos semissupervisionados que tinham suas quantidades de exemplos iniciais reduzidos para treinamento dos classificadores iniciais, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

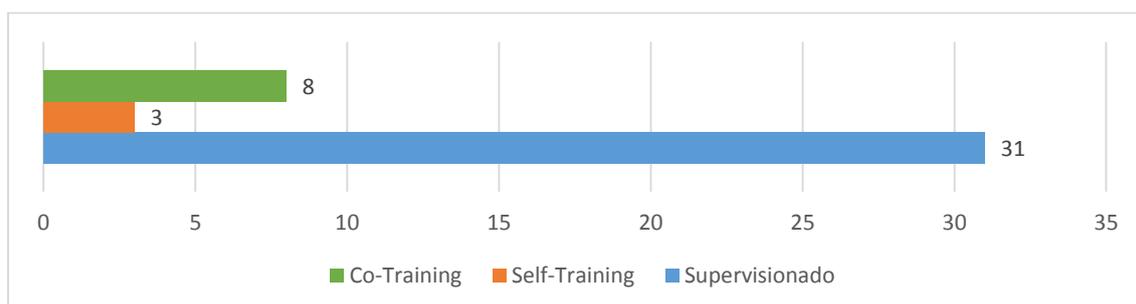


Figura 5.8: Métodos Semissupervisionado Multivisão *versus* Semissupervisionado Visão Simples *versus* Supervisionado no Nível 1 da Hierarquia

5.2. MÉTODOS HIERÁRQUICOS MULTIRRÓTULO: SEGUNDO NÍVEL

Nesta seção, são apresentados os resultados dos métodos de classificação hierárquica multirrótulo supervisionada e semissupervisionadas. É importante enfatizar que todos os resultados desta seção referem-se aos resultados obtidos no segundo nível da hierarquia. Inicialmente, serão apresentados os resultados dos métodos de classificação hierárquica multirrótulo supervisionada sob a perspectiva das técnicas multirrótulo no contexto da classificação hierárquica. Após isso, serão apresentados os resultados dos métodos de classificação hierárquica multirrótulo semissupervisionadas sob a perspectiva das técnicas multirrótulo no contexto da classificação hierárquica, do percentual de dados rotulados iniciais e, após isso, da perspectiva do percentual de incorporação do iteração. O objetivo da análise sob várias perspectivas distintas reside no estudo do impacto da escolha dos valores dos parâmetros no comportamento dos classificadores. Ao realizar esta análise, espera-se ter um indicativo de quais valores dos parâmetros apresentam os melhores resultados, considerando o universo estudado. Por fim, será realizada uma análise comparativa dos resultados obtidos pelos métodos de classificação hierárquica multirrótulo supervisionada e dos métodos de classificação hierárquica multirrótulo semissupervisionadas.

5.2.1. Melhores Resultados dos Métodos BR, CC e LP no Contexto Supervisionado

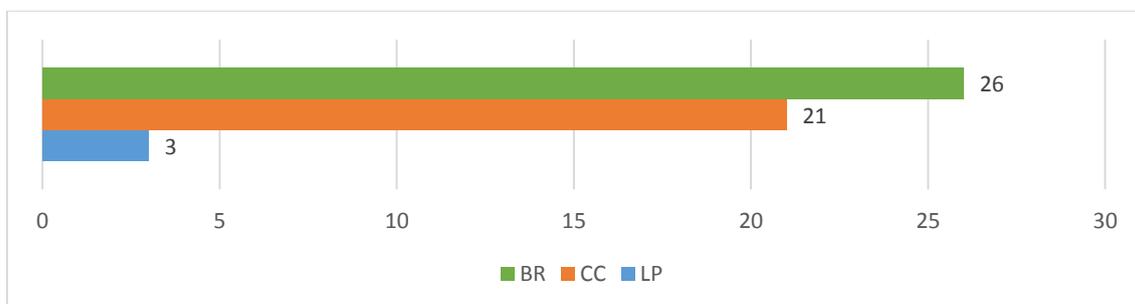


Figura 5.9: Métodos de Classificação Hierárquica Multirrótulo Supervisionados no Nível 2 da Hierarquia

A Figura 5.9 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo supervisionado no segundo nível da hierarquia, resultados obtidos após uma análise na Tabela A.9, da qual foi possível observar que os experimentos utilizando a classificação hierárquica multirrótulo supervisionada com os métodos BR, CC e LP, o método BR apresentou melhor desempenho em 53,06% dos casos (26/49), seguido do

CC com 42,86% dos casos (21/49) e os LP com 6,12% dos casos (3/49), este tendo um caso empatados com o BR. Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 81,63% dos casos (40/49). O BR mostrou-se superior do ponto de vista estatística em 48,98% dos casos (24/49), seguido pelo CC com 28,57% dos casos (14/49) e LP com 6,12% dos casos (3/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 52,50% dos casos (42/80) o Teste de Wilcoxon encontrou relevância estatística. Ao analisar a Figura 5.9, vê-se que, neste caso, o método BR obteve o melhor desempenho entre os métodos utilizados, este é um dos métodos mais utilizados e o fato deste não considerar as relações entre as classes não atrapalhou o seu desempenho, contudo, pode-se verificar relevância estatística na maioria dos casos.

5.2.2. Melhores Resultados dos Métodos BR, CC e LP no Contexto Semissupervisionado Visão Simples

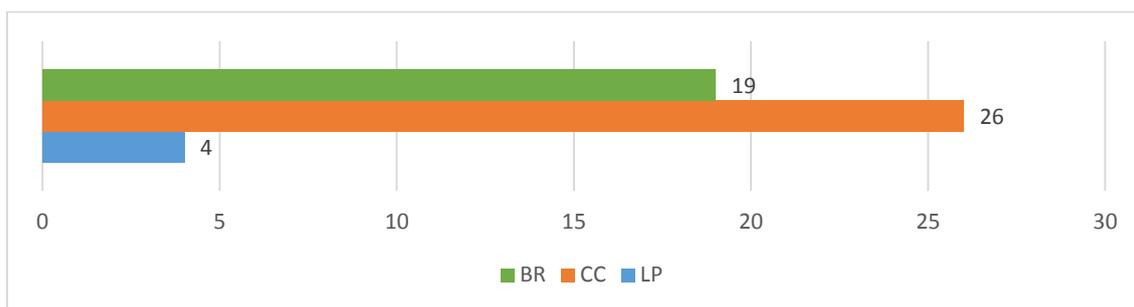


Figura 5.10: Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Visão Simples no Nível 2 da Hierarquia

A Figura 5.10 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo semissupervisionado visão simples no segundo nível da hierarquia, resultados obtidos após uma investigação na Tabela A.10, da qual foi possível observar que os experimentos utilizando a classificação hierárquica multirrótulo semissupervisionada visão simples com os métodos BR, CC e LP, o método CC apresentou melhor desempenho em 53,06% dos casos (26/49), seguido do BR com 38,78% dos casos (19/49) e o LP com 8,16% dos casos (4/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 75,51% dos casos (37/49). O CC mostrou-se superior do ponto de vista estatística em

38,78% dos casos (19/49), seguido pelo BR com 34,69% dos casos (17/49) e LP com 2,04% dos casos (1/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 24,32% dos casos (18/74) o Teste de Wilcoxon encontrou relevância estatística. Ao analisar a Figura 5.10, vê-se que, neste caso, o método CC obteve o melhor desempenho entre os métodos utilizados, isso se deve ao fato do CC considerar as relações entre as classes e as mesmas tenham ajudado na melhoria da geração do classificador, contudo, pode-se verificar relevância estatística na maioria dos casos.

5.2.3. Melhores Resultados dos Métodos BR, CC e LP no Contexto Semissupervisionado Multivisão

Verificando a Tabela A.11, é possível observar que os experimentos utilizando a classificação hierárquica multirrótulo semissupervisionada multivisão com os métodos BR, CC e LP, o método CC apresentou melhor desempenho em 36,73% dos casos (18/49), seguido do BR com 34,69% dos casos (17/49) e o LP com 28,57% dos casos (14/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 67,35% dos casos (33/49). O BR mostrou-se superior do ponto de vista estatística em 28,57% dos casos (14/49), seguido pelo CC com 20,41% dos casos (10/49) e LP com 18,37% dos casos (9/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 19,70% dos casos (13/66) o Teste de Wilcoxon encontrou relevância estatística. A Figura 5.11 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo semissupervisionado multivisão no segundo nível da hierarquia, ao analisar a mesma, vê-se que, neste caso, o método CC obteve o melhor desempenho entre os métodos utilizados, isso se deve ao fato do CC considerar as relações entre as classes e as mesmas tenham ajudado na melhoria da geração do classificador, contudo, pode-se verificar relevância estatística na maioria dos casos.

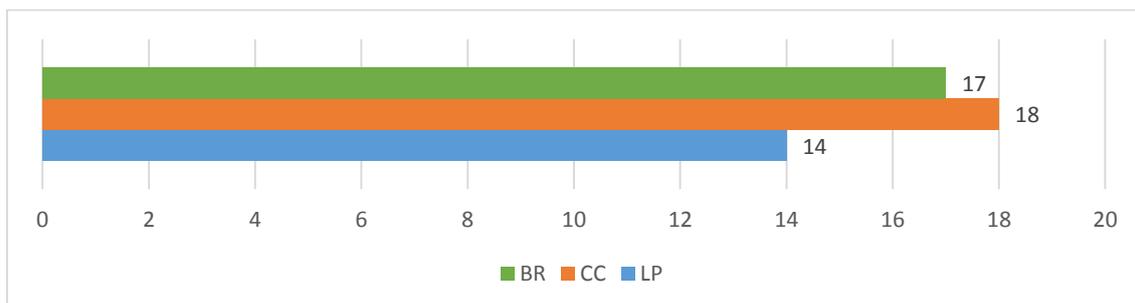


Figura 5.11: Métodos de Classificação Hierárquica Multirrotulo Semissupervisionado Multivisão no Nível 2 da Hierarquia

5.2.4. Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados

Analisando a Tabela A.12, é possível observarmos que o valor de 50% de dados não rotulados iniciais foi superior em 71,43% dos casos (35/49), enquanto que o 75% de não rotulados foi superior em 28,57% dos casos (14/49). Ao aplicar o Teste de Wilcoxon, verificou-se relevância estatística em 12,24% dos casos (6/49), neste caso, o valor de 50% foi superior estatisticamente em todos estes casos em relação ao valor de 75%. Com isso, pode-se afirmar que quanto maior o número de dados rotulados iniciais, melhor o desempenho dos classificadores, porém, não se pode comprovar a superioridade estatística na maioria dos casos. A Figura 5.12 ilustra o desempenho do método semissupervisionado visão simples para percentual inicial de exemplos não rotulados no segundo nível da hierarquia.

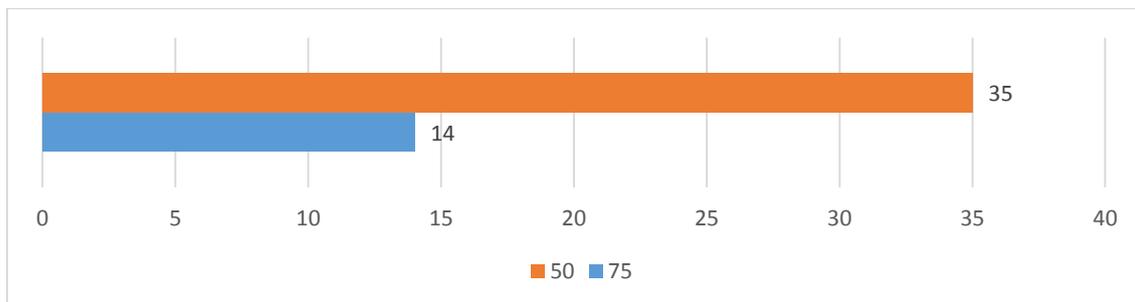


Figura 5.12: Método Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados no Nível 2 da Hierarquia

5.2.5. Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados

Explorando a Tabela A.13, é possível observarmos que o valor de 50% de dados não rotulados iniciais foi superior em 89,80% dos casos (44/49), enquanto que o 75% de não rotulados foi superior em 10,20% dos casos (5/49). Ao aplicar o Teste de Wilcoxon, verificou-se relevância estatística em 12,24% dos casos (6/49), neste caso, o valor de 50% foi superior estatisticamente em todos estes casos em relação ao valor de 75%. Com isso,

pode-se afirmar que quanto maior o número de dados rotulados iniciais, melhor o desempenho dos classificadores, porém, não se pode comprovar a superioridade estatística na maioria dos casos. A Figura 5.13 ilustra o desempenho do método semissupervisionado multivisão para percentual inicial de exemplos não rotulados no segundo nível da hierarquia.

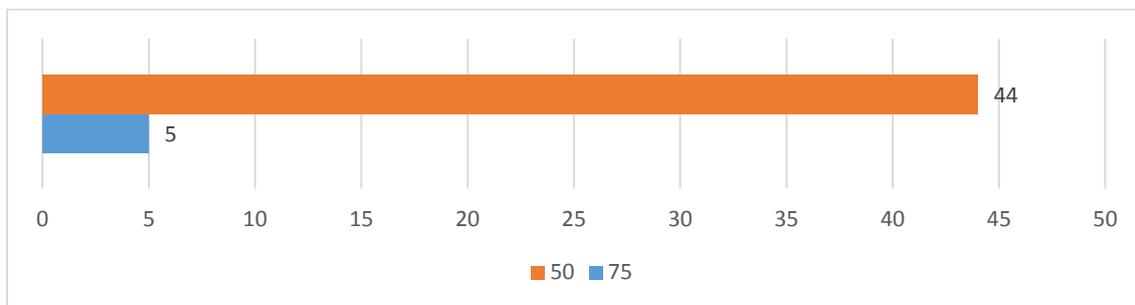


Figura 5.13: Método Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados no Nível 2 da Hierarquia

5.2.6. Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados

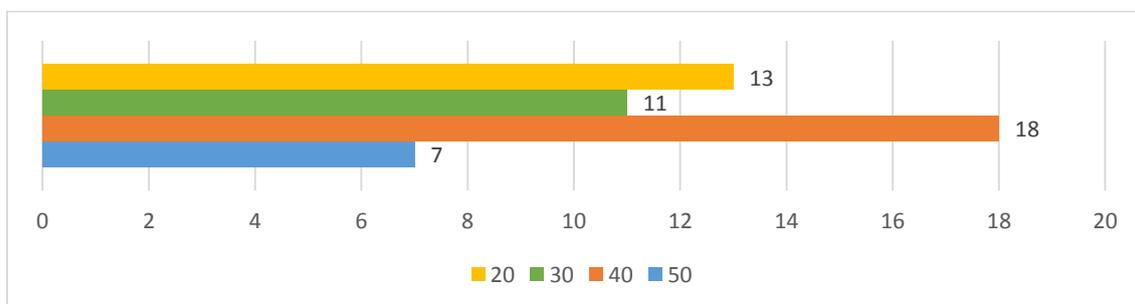


Figura 5.14: Método Semissupervisionado Visão Simples Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 2 da Hierarquia

A Figura 5.14 ilustra o desempenho do método semissupervisionado visão simples para percentual de incorporação por iteração de novos exemplos rotulados no segundo nível da hierarquia, resultados obtidos após uma investigação na Tabela A.14, da qual observa-se que em 36,73% dos casos (18/49) o valor de 40% ficou com o melhor resultado, seguido do percentual de 20% com 26,53% dos casos (13/49), o 30% com 22,45% dos casos (11/49) e o de 50% com 14,29% dos casos (7/49). Ao utilizar o Teste

de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 10,20% dos casos (5/49). O percentual de 30% mostrou-se superior do ponto de vista estatística em 6,12% dos casos (3/49), seguido pelo 40% com 4,08% dos casos (2/49) e as porcentagens de 20 e 50% com nenhum caso. Ao observar a Figura 5.14, vê-se que um dos maiores percentuais de iteração por ciclo obteve o melhor desempenho para a geração do classificador, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

5.2.7. Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados

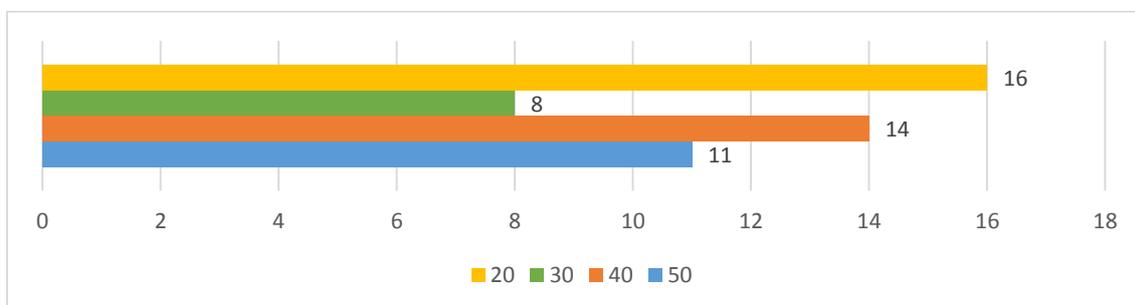


Figura 5.15: Método Semissupervisionado Multivisão Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 2 da Hierarquia

A Figura 5.15 ilustra o desempenho do método semissupervisionado multivisão para percentual incorporação por iteração de novos exemplos rotulados no segundo nível da hierarquia, resultado após averiguar a Tabela A.15, que possui os valores das porcentagens de incorporação por iteração, da qual observa-se que em 32,65% dos casos (16/49) o valor de 20% ficou com o melhor resultado, seguido do percentual de 40% com 28,57% dos casos (14/49), o 50% com 22,45% dos casos (11/49) e o de 30% com 16,33% dos casos (8/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 4,08% dos casos (2/49). Os percentuais de 20 e 40% mostraram-se superiores do ponto de vista estatística em 2,04% dos casos (1/49) para cada uma das porcentagens e as porcentagens de 30 e 50% com nenhum caso. Ao observar a Figura 5.15, vê-se que o menor percentual de iteração por ciclo obteve o melhor desempenho para a geração do classificador, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

5.2.8. Abordagem Semissupervisionada Multivisão *versus* Semissupervisionada Visão Simples *versus* Supervisionada

Explorando a Tabela A.16, observa-se que o método supervisionado foi o melhor com 40,82% dos casos (20/49), seguido do semissupervisionado Self-Training com 30,61% dos casos (15/49) e o semissupervisionado Co-Training com 28,57% dos casos (14/49). Ao aplicar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 77,55% dos casos (38/49). O supervisionado mostrou-se superior do ponto de vista estatística em 32,65% dos casos (16/49), seguido pelo semissupervisionado Co-Training com 28,57% dos casos (14/49) e o semissupervisionado Self-Training com 16,33% dos casos (8/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 34,21% dos casos (26/76) o Teste de Wilcoxon encontrou relevância estatística. A Figura 5.16 ilustra o desempenho dos métodos semissupervisionado multivisão *versus* semissupervisionado visão simples *versus* supervisionado no segundo nível da hierarquia. Ao analisar a Figura 5.16, pode-se observar que o aprendizado supervisionado foi superior que o métodos semissupervisionados, tal superioridade é oriunda do fato de que o método de aprendizado supervisionado realizava a fase de treinamento com todos os exemplos disponíveis, gerando, assim, classificadores com melhor desempenho comparados com os dos métodos semissupervisionados que tinham suas quantidades de exemplos iniciais reduzidos para treinamento dos classificadores iniciais, contudo, se pode comprovar a superioridade estatística na maioria dos casos.

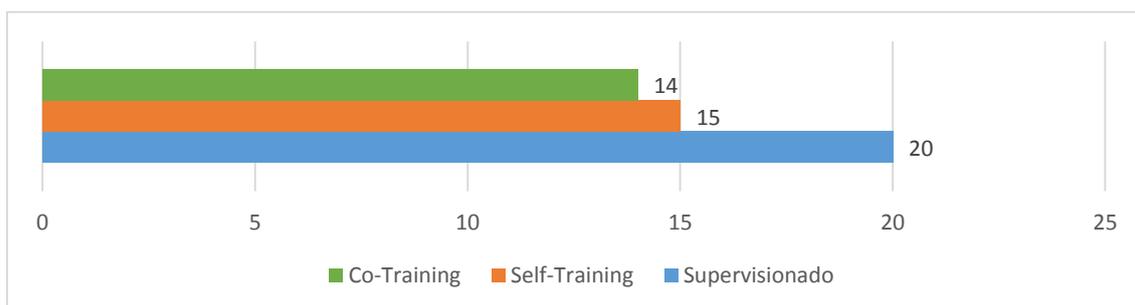


Figura 5.16: Métodos Semissupervisionado Multivisão *versus* Semissupervisionado Visão Simples *versus* Supervisionado no Nível 2 da Hierarquia

5.3. MÉTODOS HIERÁRQUICOS MULTIRRÓTULO: TERCEIRO NÍVEL

Nesta seção, são apresentados os resultados dos métodos de classificação hierárquica multirrótulo supervisionada e semissupervisionadas. É importante enfatizar que todos os resultados desta seção referem-se aos resultados obtidos no terceiro nível da hierarquia. Inicialmente, serão apresentados os resultados dos métodos de classificação hierárquica multirrótulo supervisionada sob a perspectiva das técnicas multirrótulo no contexto da classificação hierárquica. Após isso, serão apresentados os resultados dos métodos de classificação hierárquica multirrótulo semissupervisionadas sob a perspectiva das técnicas multirrótulo no contexto da classificação hierárquica, do percentual de dados rotulados iniciais e, após isso, da perspectiva do percentual de incorporação do iteração. O objetivo da análise sob várias perspectivas distintas reside no estudo do impacto da escolha dos valores dos parâmetros no comportamento dos classificadores. Ao realizar esta análise, espera-se ter um indicativo de quais valores dos parâmetros apresentam os melhores resultados, considerando o universo estudado. Por fim, será realizada uma análise comparativa dos resultados obtidos pelos métodos de classificação hierárquica multirrótulo supervisionada e dos métodos de classificação hierárquica multirrótulo semissupervisionadas.

5.3.1. Melhores Resultados dos Métodos BR, CC e LP no Contexto Supervisionado

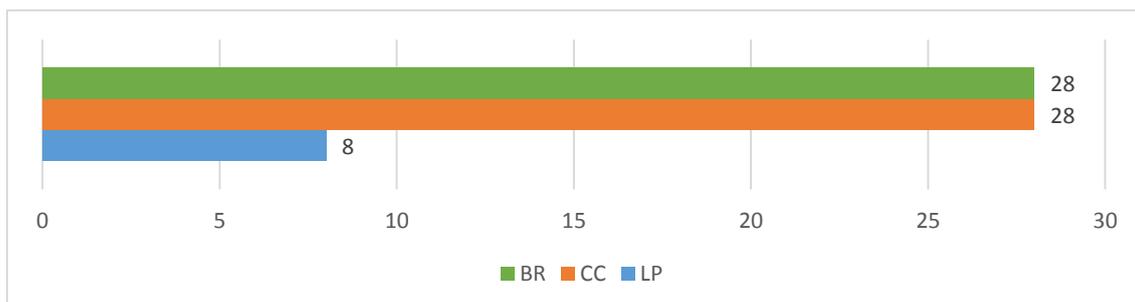


Figura 5.17: Métodos de Classificação Hierárquica Multirrótulo Supervisionados no Nível 3 da Hierarquia

A Figura 5.17 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo supervisionado no terceiro nível da hierarquia, resultados obtidos após uma análise na Tabela A.17, da qual foi possível observar que os experimentos utilizando a

classificação hierárquica multirrótulo supervisionada com os métodos BR, CC e LP, os métodos BR e CC apresentaram melhores desempenhos, empatando, em 57,14% dos casos (28/49) e ambos possuindo os melhores resultados em várias métricas de avaliação, seguindo-os, tem-se o método LP com 8,16% dos casos (8/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 71,43% dos casos (35/49). O BR mostrou-se superior do ponto de vista estatística em 55,10% dos casos (27/49), seguido pelo CC com 36,73% dos casos (18/49) e LP com 2,04% dos casos (1/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, no método BR, foi constatado que em 57,14% dos casos (40/70) o Teste de Wilcoxon encontrou relevância estatística e, no método CC, foi constatado que em 30,00% dos casos (21/70) o Teste de Wilcoxon encontrou relevância estatística. Ao analisar a Figura 5.17, vê-se que, neste caso, os métodos CC e BR obtiveram o melhor desempenho entre os métodos utilizados, isso se deve ao fato do CC ser baseado no BR, mas considerar as relações entre as classes, porém, este fato não influenciou no desempenho do classificador para que o CC obtivesse superioridade, contudo, pode-se verificar relevância estatística na maioria dos casos.

5.3.2. Melhores Resultados dos Métodos BR, CC e LP no Contexto Semissupervisionado Visão Simples

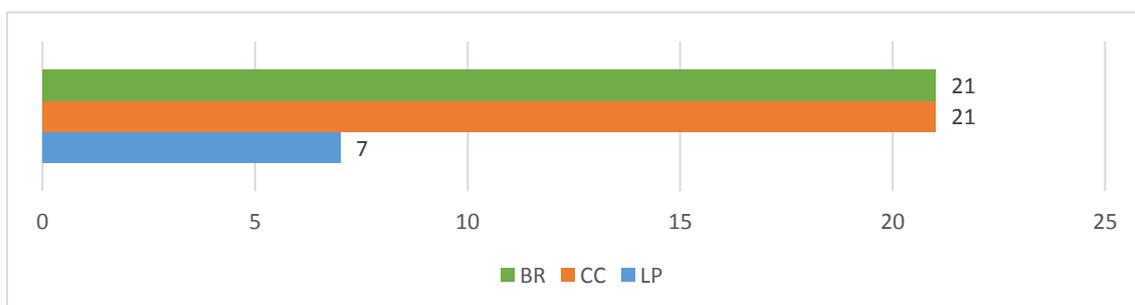


Figura 5.18: Métodos de Classificação Hierárquica Multirrótulo Semissupervisionado Visão Simples no Nível 3 da Hierarquia

A Figura 5.18 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo semissupervisionado visão simples no terceiro nível da hierarquia, resultados obtidos após uma verificação na Tabela A.18, da qual foi possível observar que os experimentos utilizando a classificação hierárquica multirrótulo semissupervisionada

visão simples com os métodos BR, CC e LP, os métodos BR e CC apresentaram melhores desempenhos, empatando, em 42,86% dos casos (21/49) e o LP com 14,29% dos casos (7/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 73,47% dos casos (36/49). O BR mostrou-se superior do ponto de vista estatística em 38,78% dos casos (19/49), seguido pelo CC com 26,53% dos casos (13/49) e LP com 8,16% dos casos (4/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, no método BR, foi constatado que em 38,89% dos casos (28/72) o Teste de Wilcoxon encontrou relevância estatística e, no método CC, foi constatado que em 19,44% dos casos (14/72) o Teste de Wilcoxon encontrou relevância estatística. Ao analisar a Figura 5.18, vê-se que, neste caso, os métodos CC e BR obtiveram o melhor desempenho entre os métodos utilizados, isso se deve ao fato do CC ser baseado no BR, mas considerar as relações entre as classes, porém, este fato não influenciou no desempenho do classificador para que o CC obtivesse superioridade, contudo, pode-se verificar relevância estatística na maioria dos casos.

5.3.3. Melhores Resultados dos Métodos BR, CC e LP no Contexto Semissupervisionado Multivisão

Investigando a Tabela A.19, é possível observar que os experimentos utilizando a classificação hierárquica multirrótulo semissupervisionada multivisão com os métodos BR, CC e LP, o método CC apresentou melhor desempenho em 36,73% dos casos (18/49), seguido do BR com 32,65% dos casos (16/49) e o LP com 30,61% dos casos (15/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 69,39% dos casos (34/49). O BR mostrou-se superior do ponto de vista estatística em 28,57% dos casos (14/49), seguido pelo CC e LP, ambos com 20,41% dos casos (10/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 17,65% dos casos (12/68) o Teste de Wilcoxon encontrou relevância estatística. A Figura 5.19 ilustra o desempenho dos métodos de classificação hierárquica multirrótulo semissupervisionado multivisão no terceiro nível da hierarquia, ao analisar a mesma, vê-se que, neste caso, o método CC obteve o melhor desempenho entre os métodos utilizados, isso se deve ao fato do CC considerar as relações entre as classes e as mesmas tenham ajudado na melhoria da geração do classificador, contudo, pode-se verificar relevância estatística na maioria dos casos.

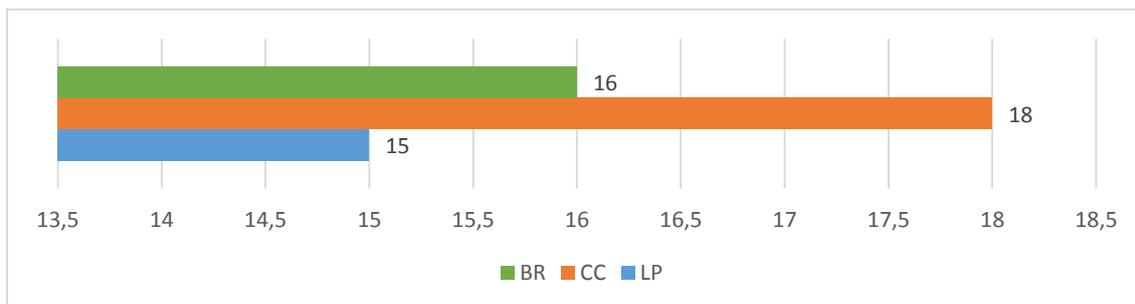


Figura 5.19: Métodos de Classificação Hierárquica Multirrotulo Semissupervisionado Multivisão no Nível 3 da Hierarquia

5.3.4. Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados

Analisando a Tabela A.20, é possível observarmos que o valor de 50% de dados não rotulados iniciais foi superior em 69,39% dos casos (34/49), enquanto que o 75% de não rotulados, ou seja, apenas 25% dos dados rotulados iniciais, obteve 30,61% dos casos (15/49). Ao realizar a os testes estatísticos, não foi verificada a relevância estatística em nenhum dos casos. Com isso, podemos afirmar que os valores utilizados não apresentam impacto relevante nos resultados apresentados, porém, o valor de 50% obteve o melhor desempenho na maioria dos casos. A Figura 5.20 ilustra o desempenho do método semissupervisionado visão simples para percentual inicial de exemplos não rotulados no terceiro nível da hierarquia.

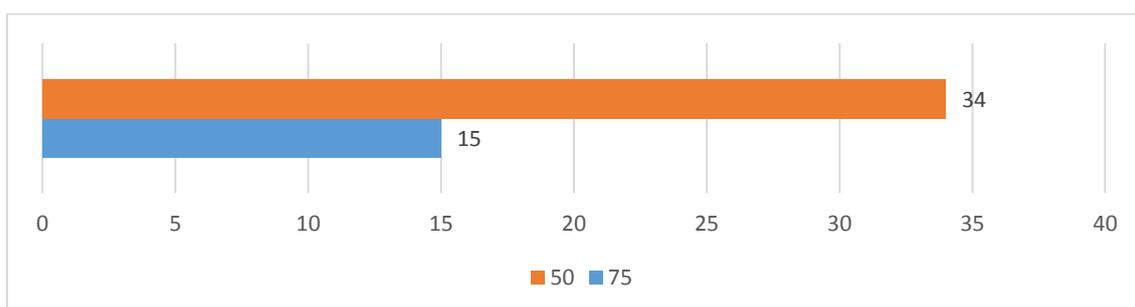


Figura 5.20: Método Semissupervisionado Visão Simples Para Percentual Inicial de Exemplos Não Rotulados no Nível 3 da Hierarquia

5.3.5. Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Rotulados

Verificando a Tabela A.21, é possível observarmos que o valor de 50% de dados não rotulados iniciais foi superior em 81,63% dos casos (40/49), enquanto que o 75% de não rotulados foi superior em 18,37% dos casos (9/49). Ao aplicar o Teste de Wilcoxon, verificou-se relevância estatística em 20,41% dos casos (10/49). A Figura 5.21 ilustra o desempenho do método semissupervisionado multivisão para percentual inicial de exemplos não rotulados no terceiro nível da hierarquia. Ao analisar a Figura 5.21, observa-se que o valor de 50% obteve melhor desempenho, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

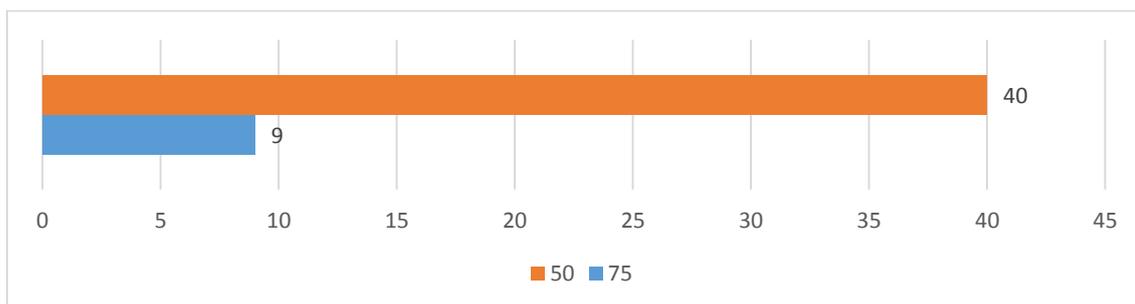


Figura 5.21: Método Semissupervisionado Multivisão Para Percentual Inicial de Exemplos Não Rotulados no Nível 3 da Hierarquia

5.3.6. Melhores Resultados no Contexto Semissupervisionado Visão Simples Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados

Investigando a Tabela A.22, com os valores das porcentagens de incorporação por iteração, observa-se que em 40,82% dos casos (20/49) o valor de 50% ficou com o melhor resultado, seguido do percentual de 30 e 40% com 20,41% dos casos (10/49) e o de 20% com 18,37% dos casos (9/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 4,08% dos casos (2/49). O percentual de 50% mostrou-se superior do ponto de vista estatística em 4,08% dos casos (2/49) e as porcentagens de 20, 30 e 40% com nenhum caso. A Figura 5.22 ilustra o desempenho do método semissupervisionado visão simples para percentual incorporação por iteração de novos exemplos rotulados no terceiro nível da hierarquia. Ao analisar a Figura 5.22, observa-se que o maior percentual de iteração por ciclo obteve o melhor desempenho para a geração do classificador, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

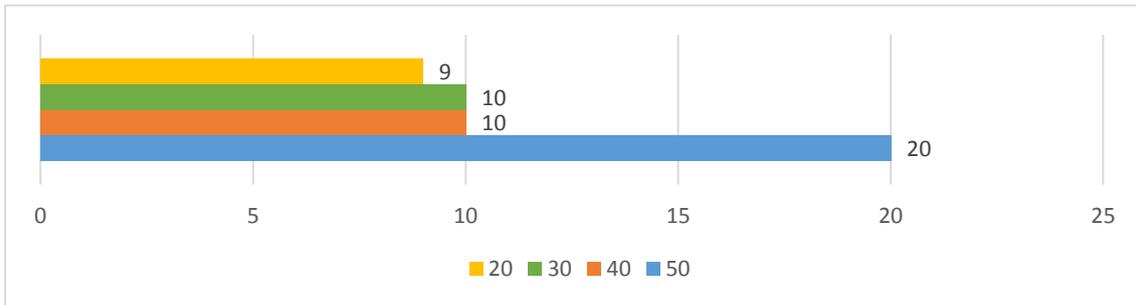


Figura 5.22: Método Semissupervisionado Visão Simples Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 3 da Hierarquia

5.3.7. Melhores Resultados no Contexto Semissupervisionado Multivisão Para Percentual de Incorporação Por Iteração de Novos Exemplos Rotulados

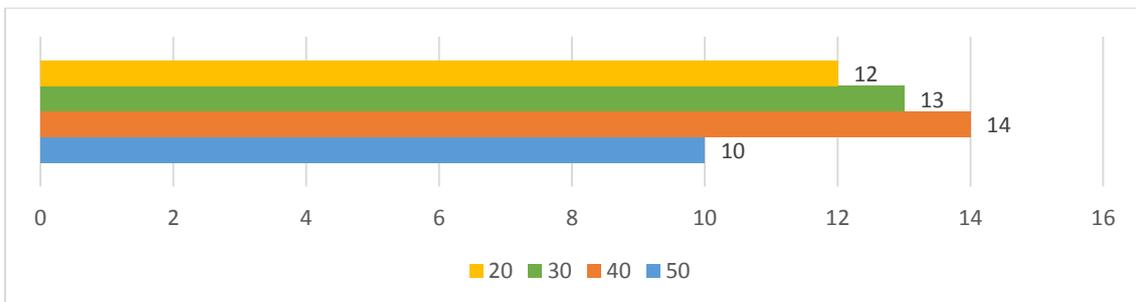


Figura 5.23: Método Semissupervisionado Multivisão Para Percentual Incorporação Por Iteração de Novos Exemplos Rotulados no Nível 3 da Hierarquia

A Figura 5.23 ilustra o desempenho do método semissupervisionado multivisão para percentual de incorporação por iteração de novos exemplos rotulados no terceiro nível da hierarquia, resultado após averiguar a Tabela A.23, que possui os valores das porcentagens de incorporação por iteração, da qual observa-se que em 28,57% dos casos (14/49) o valor de 40% ficou com o melhor resultado, seguido do percentual de 30% com 26,53% dos casos (13/49), o 20% com 24,49% dos casos (12/49) e o de 50% com 20,41% dos casos (10/49). Ao utilizar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 8,16% dos casos (4/49). O percentual de 30% mostrou-se superior do ponto de vista estatístico em 6,12% dos casos (3/49), seguido pelo 40% com 2,04% dos casos (1/49) e as porcentagens de 20 e 50% com nenhum caso. Ao alisar a Figura 5.23, observa-se que um dos maiores percentuais

de iteração por ciclo obteve o melhor desempenho para a geração do classificador, porém, não se pode comprovar a superioridade estatística na maioria dos casos.

5.3.8. Abordagem Semissupervisionada Multivisão *versus* Semissupervisionada Visão Simples *versus* Supervisionada

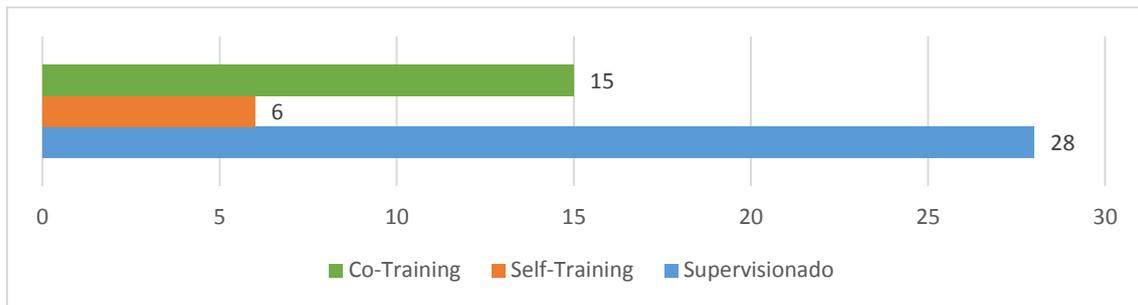


Figura 5.24: Métodos Semissupervisionado Multivisão *versus* Semissupervisionado Visão Simples *versus* Supervisionado no Nível 3 da Hierarquia

A Figura 5.24 ilustra o desempenho dos métodos semissupervisionado multivisão *versus* semissupervisionado visão simples *versus* supervisionado no terceiro nível da hierarquia, resultado após a explorar a Tabela A.24, da qual observa-se que o método supervisionado foi o melhor com 57,14% dos casos (28/49), seguido do semissupervisionado Co-Training com 30,61% dos casos (15/49) e o semissupervisionado Self-Training com 12,24% dos casos (6/49). Ao aplicar o Teste de Friedman, verificou-se que a significância estatística nas diferenças apresentadas são de 77,55% dos casos (38/49). O supervisionado mostrou-se superior do ponto de vista estatística em 42,86% dos casos (21/49), seguido pelo semissupervisionado Co-Training com 30,61% dos casos (15/49) e o semissupervisionado Self-Training com 4,08% dos casos (2/49). Ao investigar em quais amostras as diferenças estatísticas são relevantes, foi constatado que em 39,47% dos casos (30/76) o Teste de Wilcoxon encontrou relevância estatística. Ao analisar a Figura 5.24, pode-se observar que o aprendizado supervisionado foi superior que o métodos semissupervisionados, tal superioridade é oriunda do fato de que o método de aprendizado supervisionado realizava a fase de treinamento com todos os exemplos disponíveis, gerando, assim, classificadores com melhor desempenho comparados com os dos métodos semissupervisionados que tinham

suas quantidades de exemplos iniciais reduzidos para treinamento dos classificadores iniciais, contudo, se pode comprovar a superioridade estatística na maioria dos casos.

5.4. DESEMPENHO ENTRE NÍVEIS

Como já citado anteriormente, nos problemas de classificação hierárquica, o desempenho da tarefa de predição de novos exemplos, em nós mais profundos da árvore, tende a ter uma piora, pois, quanto mais fundo se deseja fazer a classificação, menos exemplos se tem para uma determinada classe para que seja realizada a fase de treinamento, com isso, se resulta em um classificador menos acurado.

Após esta afirmação, nesta sessão, fez-se uma comparação entre níveis, para cada abordagem utilizada neste trabalho, seguindo a ordem do aprendizado supervisionado, seguido do semissupervisionado com o *Self-Training* e por fim, com o algoritmo semissupervisionado *Co-Training*, afim de investigar a piora ou a melhoria de desempenho das métricas de avaliação comparando-as entre os três níveis hierárquicos.

5.4.1. Desempenho do Aprendizado Supervisionado, Semissupervisionado *Self-Training* e *Co-Training*

Ao analisar as Tabelas A.25, A.26 e A.27, observa-se que apenas a métrica HLoss obteve o melhor desempenho no terceiro nível hierárquico, todas as outras métricas obtiveram melhores resultados no primeiro nível da hierarquia, vale salientar que para a HiLoss, considera-se como primeiro nível o nível dois. Com isso, vê-se que, com exceção da HLoss, todas as outras métricas de avaliação vão obtendo piores resultados ao passar para um nível mais abaixo na classificação hierárquica multirrótulo. Tal padrão foi obtido nos três métodos de aprendizado abordados neste trabalho. A Figura 5.25 ilustra o desempenho entre níveis no aprendizado supervisionado, semissupervisionado *Self-Training* e *Co-Training*.

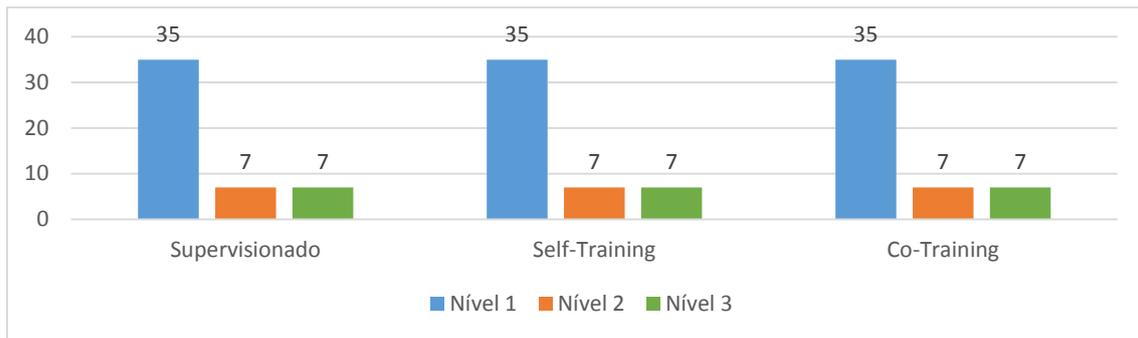


Figura 5.25: Desempenho Entre Níveis no Aprendizado Supervisionado, Semissupervisionado *Self-Training* e *Co-Training*

5.5. CONSIDERAÇÕES FINAIS

Neste capítulo foram apresentados os resultados experimentais obtidos pelos métodos de classificação hierárquica multirrótulo utilizando as abordagens de aprendizado supervisionado, semissupervisionado visão simples e semissupervisionado multivisão. Os experimentos utilizando os métodos de classificação hierárquica multirrótulo foram realizados nos três primeiros níveis da hierarquia. Para cada um dos níveis da hierarquia, os métodos de classificação hierárquica multirrótulo que utilizam a abordagem semissupervisionada foram apresentados sob duas perspectivas: percentual inicial de exemplos não rotulados e percentual de incorporação de novos exemplos rotulados por iteração. O objetivo da análise sob várias perspectivas distintas reside no estudo do impacto da escolha dos valores dos parâmetros no comportamento dos classificadores.

Ao analisar os resultados na perspectiva de percentual de exemplos não rotulados, foram analisados os dois valores seguintes: 50 e 75%. A análise dos resultados obtidos mostrou que, em geral, que a maior quantidade de exemplos rotulados obteve os melhores resultados. Isso se deve a um classificador com melhor acurácia na execução inicial do algoritmo semissupervisionado, além disso, quanto menor o percentual inicial de rotulados, maior a probabilidade de inclusão de ruído durante o processo de atribuição automática de rótulos. Do ponto de vista estatístico, não se pode comprovar superioridade estatística na maioria dos casos em nenhum dos experimentos.

Na perspectiva de percentual de novos exemplos rotulados por iteração, percebeu-se que na maioria dos casos, os melhores resultados foram obtidos com uma

maior porcentagem de rotulação por ciclo. Isso se deve a maior quantidade de novos exemplos à cada iteração, gerando um melhor classificador naquele momento. Do ponto de vista estatístico, não se pode comprovar superioridade estatística na maioria dos casos em nenhum dos experimentos.

Ainda foi realizado um estudo sob a perspectiva hierárquica multirrótulo, dos quais foram utilizados os algoritmos BR, CC e LP, todos estes aplicados nas abordagens supervisionada, semissupervisionada visão simples e semissupervisionada multivisão. No nível 1 da hierarquia, o método LP foi o melhor nas três abordagens, já no nível 2, o método CC obteve o melhor desempenho com os dois métodos semissupervisionados, ficando o supervisionado com o algoritmo BR e no nível 3, os métodos BR e CC empataram tanto na abordagem supervisionada, como na abordagem semissupervisionada visão simples, porém, o semissupervisionado multivisão obteve melhor resultado com o CC, tornando-o assim, o melhor método no terceiro nível da hierarquia. O método LP obteve o melhor desempenho entre os métodos utilizados em todos os experimentos no primeiro nível da hierarquia, isso se deve ao fato do LP considerar as relações entre as classes e a quantidade de subconjuntos de classes gerados pelo mesmo tenha ajudado na melhoria da geração do classificador, porém, o LP possui uma desvantagem quando se trabalha com uma grande quantidade de classes, o mesmo pode gerar uma grande quantidade de subconjuntos de rótulos, aumentando o custo computacional e diminuindo a quantidade de exemplos associados a uma classe, o que acarreta na diminuição da acurácia dos classificadores, tal fato fez com que o LP obtivesse os piores resultados a partir do segundo nível da hierarquia, dando o lugar de melhor método para o CC, o melhor nos níveis 2 e 3, que também considera a relação entre classes, mas não cai no problema gerado pelo LP. Do ponto de vista estatístico, no primeiro nível, verificou-se superioridade estatística em 1 de 3 experimentos, neste caso, foi o semissupervisionado visão simples. Já nos níveis 2 e 3 da hierarquia se pode verificar superioridade estatística em todos os experimentos.

Em seguida, foi realizada uma análise de uma perspectiva geral, possibilitando uma comparação entre as abordagens supervisionada, semissupervisionada visão simples e semissupervisionada multivisão nos diferentes níveis da hierarquia. Em todos os níveis da hierarquia a abordagem supervisionada foi superior, o que era de se esperar, pois o classificador era gerado com todos exemplos rotulados disponíveis nos conjuntos de dados, ao contrário dos semissupervisionados, que iniciavam apenas com uma

porcentagem desses exemplos, gerando assim, um classificador com um pior desempenho em relação ao supervisionado. Contudo, pode-se verificar relevância estatística apenas nos níveis 2 e 3.

Por fim, analisou-se as abordagens supervisionada, semissupervisionada visão simples e a multivisão, nos três níveis da classificação hierárquica multirrótulo e cada abordagem foi analisada separadamente, a fim de investigar se ao ir para o próximo nível hierárquico haveria uma piora ou melhoria do desempenho das métricas de avaliação. Como resultado, vê-se que, com exceção da HLoss, todas as outras métricas de avaliação vão obtendo piores resultados ao passar para um nível mais abaixo na classificação hierárquica multirrótulo, este padrão se repete nas três abordagens de aprendizado. Confirmando, assim, a hipótese mostrada.

Neste contexto, pode-se afirmar que, de modo geral, a utilização do aprendizado supervisionado em problemas de classificação hierárquica multirrótulo apresentou os melhores resultados em comparação com as abordagens semissupervisionadas.

Capítulo 6

CONCLUSÃO

A classificação de dados tradicional trata do aprendizado de um conjunto de exemplos em que cada exemplo é associado a somente uma única classe de um conjunto finito de classes. Embora a classificação tradicional seja bastante utilizada, há vários domínios nos quais os exemplos podem estar associados a mais de uma classe do conjunto finito de classes, sendo chamados de problemas de classificação multirrótulo. Por outro lado, há também um grande número de problemas de classificação, nos quais as classes a serem preditas estão dispostas em uma estrutura hierárquica, sendo conhecidos como problemas de classificação hierárquica. Ademais, recentemente, têm-se identificado a existência de muitos problemas de classificação hierárquica mais complexos, em que duas ou mais classes do conjunto finito de classes podem ser atribuídas simultaneamente à mesma instância. Esses problemas nos quais são combinadas a classificação hierárquica e a classificação multirrótulo são chamados de problemas de classificação hierárquica multirrótulo.

Por sua vez, o aprendizado semissupervisionado tem emergido como uma abordagem bastante interessante para o aprendizado de máquina, principalmente em problemas onde a quantidade de exemplos não rotulados gerados é alta e a quantidade de exemplos rotulados é muito baixa, ou seja, baixa dimensionalidade de dados.

Como mencionado anteriormente, o objetivo do trabalho proposto é o estudo da aplicação de técnicas de aprendizado semissupervisionado multidescrição em tarefas de classificação hierárquica multirrótulo, afim de verificar se a utilização desta traz melhorias nos resultados do processo de classificação de dados.

Com isso, um estudo investigativo foi realizado para analisar comparativamente os métodos de classificação hierárquica multirrótulo nas perspectivas supervisionada, semissupervisionada visão simples e semissupervisionada multivisão com as seguintes configurações:

- Método BR com aprendizado supervisionado, aprendizado semissupervisionado com o *Self-Training* e o semissupervisionado *Co-Training*.
- Método CC com aprendizado supervisionado, aprendizado semissupervisionado com o *Self-Training* e o semissupervisionado *Co-Training*.
- Método LP com aprendizado supervisionado, aprendizado semissupervisionado com o *Self-Training* e o semissupervisionado *Co-Training*.

A análise do comportamento dos experimentos foram realizadas nos três níveis dos métodos de classificação hierárquica multirrótulo. Além disso, as abordagens semissupervisionadas foram analisadas nas duas perspectivas: percentual inicial de exemplos não rotulados e percentual de incorporação de novos exemplos rotulados por iteração. Foi realizada uma análise comparativa dos métodos de classificação supervisionada e as semissupervisionadas, sendo a supervisionada a abordagem que obteve os melhores resultados em todos os níveis. Contudo, as abordagens semissupervisionadas ainda podem vir a ser interessantes para problemas de classificação como um todo, além disso, pode ser a solução para uma baixa dimensionalidade de dados, pois, tem-se muito para ser pesquisado e conseqüentemente gerar trabalhos com contribuições exclusivas como este. Outro campo que tem muito a ser explorado é a classificação hierárquica multirrótulo, pois ainda existem poucas métricas disponíveis para serem utilizadas.

Por fim, foi feita uma investigação de desempenho por nível de cada tipo de aprendizado utilizado nos problemas de classificação hierárquica multirrótulo, vemos que, com exceção da HLoss, todas as outras métricas de avaliação vão obtendo piores resultados ao passar para um nível mais abaixo na classificação hierárquica multirrótulo.

Tal fato ocorreu tanto no aprendizado supervisionado quanto nas duas abordagens semissupervisionadas. Confirmando a hipótese apresentada.

Desse modo, este trabalho trouxe contribuições para pesquisas nas áreas de classificação hierárquica multirrótulo e aprendizado semissupervisionado, utilizando vários algoritmos combinando estas áreas. Os resultados utilizados podem servir como guia para novas pesquisas, gerando, assim, novos trabalhos.

6.1. TRABALHOS FUTUROS

A seguir são apresentados alguns pontos que podem ser desenvolvidos em futuras pesquisas envolvendo os métodos utilizados:

- Utilizar bases de dados maiores, com mais instâncias, atributos e rótulos, tendo em vista que neste trabalho foram utilizadas bases de dados pequenas;
- Realizar este mesmo estudo em níveis mais profundos da hierarquia, sendo que neste foram utilizados três níveis hierárquicos;
- Utilizar as abordagens supervisionada, semissupervisionada visão simples e multidescrição em problemas apenas de classificação multirrótulo;
- Propor novas medidas para avaliação da classificação hierárquica multirrótulo;
- Realizar uma investigação utilizando a classificação supervisionada com a mesma quantidade de instâncias utilizadas para a geração do classificador inicial na abordagem semissupervisionada.
- Utilizar um fator de confiança para inserção no conjunto de novos exemplos rotulado e de seleção de melhores exemplos no conjunto de dados para gerar os classificadores no aprendizado semissupervisionado, objetivando investigar se há melhoria no desempenho da classificação de dados com baixa dimensionalidade de exemplos no conjunto de dados.

REFERÊNCIAS

ABNEY, S. P. **Semisupervised Learning for Computational Linguistics**. Michigan, AA, USA: Chapman & Hall, p. 320, 2008.

BLUM, A.; MITCHELL, T. **Combining Labeled and Unlabeled Data with Co-training**. New York, NY, USA: Morgan Kaufmann Publishers, pp. 91-100, 1998.

BOUTELL, M. R.; LUO, J.; SHEN, X.; BROWN, C. M. **Learning Multi-label Scene Classification**. *Pattern Recognition*, v. 37, n. 9, p. 1757-1771, 2004.

BRAGA, I. A. **Aprendizado Semissupervisionado Multidescrição em Classificação de Textos**. Dissertação (Mestrado) – Universidade de São Paulo (USP), São Carlos, 2010.

CALDAS, W, L. **Proposta de Dois Métodos Semi-Supervisionados Baseados na Máquina de Aprendizagem Mínima Utilizando Co-Training**. Dissertação (Mestrado) – Universidade Federal do Ceará (UFC), Fortaleza, 2017.

CERRI, R. **Técnicas de Classificação Hierárquica Multirrótulo**. Dissertação (Mestrado) – Universidade de São Paulo (USP), São Carlos, 2010.

CERRI, R.; CARVALHO, A. C. P. L. F. de. **Comparing Local and Global Hierarchical Multilabel Classification Methods Using Decision Trees**. In: *Proceedings of the Brazilian Symposium on Bioinformatics, Lecture Notes on Bioinformatics*. Porto Alegre, Brazil: Springer-Verlag, p.109-120, 2009.

CERRI, R.; BARROS, R. C.; CARVALHO, A. C. P. L. F. **Redes Neurais Artificiais Aplicadas a Problemas de Classificação Hierárquica Multirrótulo**. X Congresso Brasileiro de Inteligência Computacional (CBIC'2011). Fortaleza, CE, 2011.

CESA-BIANCHI, N.; GENTILE, C.; ZANIBONI, L. **Hierarchical Classification: Combining Bayes With SVM**. In: *Proceedings of the 23rd international conference on Machine learning*. New York, NY, USA: ACM (ICML 06), p. 177_184, 2006.

CHERMAN, E. A. **Aprendizado de Máquina Multirrótulo: Explorando a Dependência de Rótulos e o Aprendizado Ativo**. Tese (Doutorado) – Universidade de São Paulo (USP), São Carlos, 2013.

CHU, S.; DERISI, J.; EISEN, M.; MULHOLL, J.; BOTSTEIN, D.; BROWN, P. O.; HERSKOWITZ, I. **The Transcriptional Program of Sporulation in Budding Yeast.** *Science*, v. 282, p. 699-705, 1998.

CLARE, A. **Machine Learning and Data Mining for Yeast Functional Genomics.** Tese (Doutorado) - University of Wales, Aberystwyth, 2003.

CLARE, A.; KING, R. D. **Knowledge Discovery in Multi-label Phenotype Data.** In: RAEDT, L. D.; SIEBES, A. (Ed.). *PKDD*. Freiburg, Germany: Springer, (Lecture Notes in Computer Science, v. 2168), p. 42-53, 2001.

COSTA, E. **Investigação de Técnicas de Classificação Hierárquica Para Problemas de Bioinformática.** Dissertação (Mestrado) – Universidade de São Paulo (USP), São Carlos, 2008.

DERISI, J. L.; IYER, V. R.; BROWN, P. O. **Exploring the Metabolic and Genetic Control of Gene Expression on a Genomic Scale.** *Science*, AMER ASSOC ADVANCEMENT SCIENCE, v. 278, n. 5338, p. 680-686, 1997.

DIAS, T. N. **Desenvolvimento de Técnicas de Seleção de Atributos no Contexto da Classificação Hierárquica Multirrótulo.** Dissertação (Mestrado) – Universidade Federal de Ouro Preto (UFOP), Ouro Preto, 2015.

DIPLARIS, S.; TSOUMAKAS, G.; MITKAS, P.; VLAHAVAS, I. **Protein Classification with Multiple Algorithms.** In: *Proceedings of the 10th Panhellenic Conference on Informatics (PCI 2005)*. Volos, Greece: Springer-Verlag, p. 448-456, 2005.

DU, J.; LING, C. X.; ZHOU, Z.-H. **When does Co-training Work in Real Data?** *IEEE Transactions on Knowledge and Data Engineering* 23(5): 788 – 799. Setembro 2010.

EISEN, M. B.; SPELLMAN, P. T.; BROWN, P. O.; BOTSTEIN, D. **Cluster Analysis and Display of Genome-wide Expression Patterns.** *Proceedings of the National Academy of Sciences*, National Academy of Sciences, Stanford, CA, USA, v. 95, n. 25, p. 14863-14868, dez. 1998.

FACELLI, K.; LORENA, A. C.; GAMA, J.; CARVALHO, A. C. P. d. L. F. d. **Inteligência Artificial: Uma Abordagem de Aprendizagem de Máquina.** Rio de Janeiro: LTC, 2011.

FENG, S.; FU, P.; ZHENG, W. **A Hierarchical Multi-Label Classification Algorithm for Gene Function Prediction.** *Algorithms*, v. 10, n. 138, 2017.

FREITAS, A.; CARVALHO, A. C. de. **A Tutorial on Hierarchical Classification with Applications in Bioinformatics**. In: Research and Trends in Data Mining Technologies and Applications, Idea Group, 2007. UK: Idea Group, cap. VII, 2007.

FRIEDMAN, M. **The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance**. Journal of the American Statistical Association, 32(200), 675-701, 1937.

GASCH, A. P.; HUANG, M.; METZNER, S.; BOTSTEIN, D.; ELLEDGE, S. J.; BROWN, P. O. **Genomic Expression Responses to DNA-Damaging Agentes and the Regulatory Role of the Yeast ATR Homolog Mec1p**. Molecular Biology of the Cell, v. 12, p. 2987-3003, 2001.

GASCH, A. P.; SPELLMAN, P. T.; KAO, C. M.; CARMEL-HAREL, O.; EISEN, M. B.; STORZ, G.; BOTSTEIN, D.; BROWN, P. O. **Genomic Expression Programs in the Response of Yeast Cells to Environmental Changes**. Molecular Biology of the Cell, v.11, n. 12, p. 4241-4257, 2000.

KATAKIS, I.; TSOUMAKAS, G.; VLAHAVAS, I. **Multilabel Text Classification for Automated Tag Suggestion**. In: Proceedings of the ECML/PKDD 2008 Discovery Challenge. [S.l.: s.n.], 2008.

KIRITCHENKO, S.; MATWIN, S. **Email Classification with Co-training**. In: Proceedings of the 2001 Conference of the Centre for Advanced Studies on Collaborative Research (CASCON2001). Toronto, CA: Ibm Press, p. 192 – 201, 2001.

MATSUBARA, E. T.; MONARD, M. C.; BATISTA, G. E. A. P. A. **Multi-view Semisupervised Learning: An Approach to Obtain Different Views from Text Datasets**. LAPTEC, p. 97–104, 2005.

METZ, J. **Abordagem Para Aprendizado Semissupervisionado Multirrotulo e Hierárquico**. Tese (Doutorado) – Universidade de São Paulo (USP), São Carlos, 2011.

MEWES, H.-W.; FRISHMAN, D.; GRUBER, C.; GEIER, B.; HAASE, D.; KAPS, A.; LEMCKE, K.; MANNHAUPT, G.; PFEIFFER, F.; SCHÜLLER, C. M.; STOCKER, S.; WEIL, B. **Mips: A Database for Genomes and Protein Sequences**. Nucleic Acids Research, v. 28, n. 1, p. 37-40, 2000.

MITCHELL, T. M. **Machine Learning**. New York: McGraw-Hill, 1997.

MONARD, M. C. **Sistemas Inteligentes: Fundamentos e Aplicações, Capítulo Aprendizado de Máquina**. Manole, 2003.

NASIERDING, G.; TSOUMAKAS, G.; KOUZANI, A. Z. **Clustering Based Multi-label Classification for Image Annotation and Retrieval**. In: Proc. 2009 IEEE International Conference on Systems, Man, and Cybernetics SMC'09. Texas, USA: IEEE, p. 4514-4519, 2009.

READ, J.; PFAHRINGER, B.; HOLMES, G.; FRANK, E. **Classifier Chains for Multi-label Classification**. Machine Learning, v. 85, n. 3, p. 333-359, 2011.

RODRIGUES, F. M. **Uso de Confiabilidade na Rotulação de Exemplos em Problemas de Classificação Multirrotulo com Aprendizado Semissupervisionado**. Dissertação (Mestrado) - Universidade Federal do Rio Grande do Norte (UFRN), Natal, 2014.

ROSENBERG, C.; HEBERT, M.; SCHNEIDERMAN, H. **Semi-supervised Self-training of Object Detection Models**. In: Seventh IEEE Workshop on Applications of Computer Vision. Breckenridge, CO, USA: IEEE, p. 29-36, 2005.

ROTH, F. P.; HUGHES, J. D.; ESTEP, P. W.; CHURCH, G. M. **Finding DNA Regulatory Motifs within Unaligned Noncoding Sequences Clustered by Whole-genome MRNA Quantitation**. Nature biotechnology, Nature Publishing Group, v. 16, n. 10, p. 939-945, out. 1998. ISSN 1087-0156.

SANCHES, M. K. **Aprendizado de Máquina Semi-supervisionado: Proposta de um Algoritmo para Rotular Exemplos a Partir de Poucos Exemplos Rotulados**. Dissertação (Mestrado) – Universidade de São Paulo (USP), São Carlos, 2003.

SANTOS, A. M. **Investigando a Combinação de Técnicas de Aprendizado Semissupervisionado e Classificação Hierárquica Multirrotulo**. Tese (Doutorado) - Universidade Federal do Rio Grande do Norte (UFRN), Natal, 2012.

SANTOS, A. M.; CANUTO, A. M. P. **Applying Semi-supervised Learning in Hierarchical Multi-label Classification**. Expert Systems with Applications, v. 41, p. 6075-6085, 2014. ISSN 0957-4174.

SPELLMAN, P. T.; SHERLOCK, G.; ZHANG, M. Q.; IYER, V. R.; ANDERS, K.; EISEN, M. B.; BROWN, P. O.; BOTSTEIN, D.; FUTCHER, B. **Comprehensive Identification of Cell Cycle Regulated Genes of the Yeast *Saccharomyces Cerevisiae* by Microarray Hybridization**. Molecular Biology of the Cell, American Society for Cell Biology, v. 9, n. 12, p. 3273-3297, dez. 1998.

TSOUMAKAS, G.; KATAKIS, I. **Multi Label Classification: An Overview**. International Journal of Data Warehouse and Mining, Idea Group Publishing, v. 3, n. 3, p. 1-13, 2007.

TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. **Effective and Efficient Multilabel Classification in Domains with Large Number of Labels**. Proc. ECML/PKDD 2008. Workshop on Mining Multidimensional Data (MMD'08), 2008.

TSOUMAKAS, G.; KATAKIS, I.; VLAHAVAS, I. **Mining Multi-Label Data**. In: Data Mining and Knowledge Discovery Handbook, Part 6, O. Maimon, L. Rokach (Ed.), Springer, 2nd edition, p. 667-685, 2010.

WEKA. **Data Mining Software In Java**. 2017.

WIECZORKOWSKA, A.; SYNAK, P.; RAs, Z. **Multi-label Classification of Emotions in Music**. In: Berlin, Heidelberg: Springer Berlin / Heidelberg, p. 307-315, 2006.

WILCOXON, F. **Individual Comparisons by Ranking Methods**. Biometrics bulletin, 1(6):80– 83, 1945.

XU, J.; HE, H.; MAN, H. **Dcpe Co-training for Classification**. Neurocomputing, Elsevier, v. 86, p. 75–85, 2012.

ZHANG, L.; SHAH, S. K.; KAKADIARIS, I. A. **Hierarchical Multi-Label Classification using Fully Associative Ensemble Learning**. Pattern Recognition, v. 70, p. 89-103, 2017.

ZHOU, Z.; LI, M. **Semisupervised Regression with Cotraining-style Algorithms**. Ieee Transactions on Knowledge And Data Engineering, v. 19, n. 11, p.1, 2007.

APÊNDICE A – Valores dos Experimentos Realizados

Neste apêndice são apresentados todos os valores dos experimentos para o problema de classificação hierárquica multirrótulo supervisionado e semissupervisionado realizados nesta dissertação.

Os resultados são apresentados na seguinte ordem: Melhores resultados entre os métodos multirrótulo para o aprendizado supervisionado, semissupervisionado visão simples e multivisão; Melhores resultados para percentual de exemplos não rotulados iniciais tanto para semissupervisionado visão simples quanto para a estratégia multivisão; Melhores resultados para percentual de incorporação por ciclo de novos exemplos rotulados para os dois contextos semissupervisionados utilizados; Comparação entre as três abordagens empregadas: supervisionada *versus* semissupervisionada com *Self-Training versus* semissupervisionada com *Co-Training*.

A ordem apresentada anteriormente é seguida nos três níveis da hierarquia, as Tabelas A.1 a A.8 apresentam os resultados no primeiro nível, as Tabelas A.9 a A.16 apresentam os do segundo nível e as Tabelas A.17 a A.24 exibem os resultados obtidos no terceiro nível.

Por fim, as Tabelas A.25 a A.27 possuem os resultados de desempenho obtidos entre os três níveis da classificação hierárquica multirrótulo para cada estratégia de aprendizado utilizada.

As Tabelas A.1, A.9 e A.17 pertencem aos métodos de classificação hierárquica multirrótulo supervisionada (BR, CC e LP). Já as Tabelas A.2, A.10 e A.18 pertencem aos métodos de classificação hierárquica multirrótulo semissupervisionada visão simples (BR, CC e LP). Para chegar nesses valores, inicialmente foi selecionado os melhores resultados para os valores de incorporação por ciclo, após isso, se obteve os melhores resultados para cada valor inicial de exemplos não rotulados, com isso, obtive-se os

melhores resultados por método multirrótulo. Assim, foram obtidas as Tabelas A.2, A.10 e A.18. Com essa mesma estratégia, obteve-se as Tabelas A.3, A.11 e A.19, mas neste caso, estas pertencem aos métodos de classificação hierárquica multirrótulo semissupervisionada multivisão (BR, CC e LP). As Tabelas A.1, A.2 e A.3 são do nível 1 da hierarquia, já as Tabelas A.9, A.10 e A.11 são do nível 2 da hierarquia e, por fim, as Tabelas A.17, A.18 e A.19 pertencem ao nível 3 da hierarquia. Em cada tabela, os melhores resultados apresentados em cada uma das medidas foram sombreados. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **FRI** marcados em negrito. A coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas, utilizando a seguinte nomenclatura:

- A: Método BR x CC
- B: Método BR x LP
- C: Método CC x LP

As Tabelas A.4, A.12 e A.20 apresentam a média e o desvio padrão de todos os métodos de classificação hierárquica multirrótulo semissupervisionada visão simples (BR, CC e LP) por percentual inicial de exemplos não rotulados da base de dados. Para chegar nesses valores, inicialmente foi obtido os melhores resultados nos percentuais de incorporação por ciclo, com isso, conseguiu-se os melhores resultados para cada percentual inicial de exemplos não rotulado, após isso, foi feita a junção, desse melhores resultados, de acordo com sua respectiva porcentagem em cada método multirrótulo, que por sua vez, resultou nos valores apresentados nas Tabela A.4, A.12 e A.20, que são os melhores resultados dos percentuais de exemplos não rotulados iniciais extraídos dos métodos multirrótulo. Utilizando esta mesma estratégia de coleta de resultados, obtivemos as Tabelas A.5, A.13 e A.21, porém, estas pertencem a técnica semissupervisionada multivisão. As Tabelas A.4 e A.5 são do nível 1 da hierarquia, já as Tabelas A.12 e A.13 são do nível 2 da hierarquia e, por fim, as Tabelas A.20 e A.21 pertencem ao nível 3 da hierarquia. Os experimentos realizados utilizaram dois valores para o percentual inicial de exemplos não rotulados: 50 e 75%. Em cada tabela, os melhores resultados apresentados em cada uma das medidas foram sombreados. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **WIL** marcados em negrito.

As Tabela A.6, A.14 e A.22 apresentam a média e o desvio padrão de todos os métodos de classificação hierárquica multirrótulo semissupervisionada visão simples (BR, CC e LP) por percentual de incorporação por ciclo. Para chegar nesses valores, foi feita a junção das respectivas porcentagens em cada método multirrótulo e para cada percentual de rotulados dos respectivos métodos. Desse modo, obtive-se as Tabela A.6, 14 e A.22. Com essa mesma estratégia, obteve-se as Tabelas A.7, A.15 e A.23, mas neste caso, estas pertencem aos métodos de classificação hierárquica multirrótulo semissupervisionada multivisão (BR, CC e LP). As Tabelas A.6 e A.7 são do nível 1 da hierarquia, já as Tabelas A.14 e A.15 são do nível 2 da hierarquia e, por fim, as Tabelas A.22 e A.23 pertencem ao nível 3 da hierarquia. Os experimentos realizados utilizaram dois valores para o percentual inicial de exemplos não rotulados: 20, 30, 40 e 50%. Em cada tabela, os melhores resultados apresentados em cada uma das medidas foram sombreados. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **FRI** marcados em negrito. A coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi verificada significância estatística nas diferenças apresentadas, utilizando a seguinte nomenclatura:

- A: Percentual 20 x 30
- B: Percentual 20 x 40
- C: Percentual 20 x 50
- D: Percentual 30 x 40
- E: Percentual 30 x 50
- F: Percentual 40 x 50

As Tabelas A.8, A.16 e A.24 apresentam a média e o desvio padrão de todos os métodos de classificação hierárquica multirrótulo supervisionada, semissupervisionado visão simples e multivisão (BR, CC e LP), nos níveis 1, 2 e 3 respectivamente. Essas tabelas são oriundas dos métodos multirrótulo que obtiveram o melhor resultado nas abordagens supervisionada, semissupervisionada visão simples e semissupervisionada multivisão em cada nível da hierarquia. Os melhores resultados apresentados em cada uma das medidas foram sombreados. Os resultados nos quais os testes estatísticos mostraram relevância do ponto de vista estatístico tiveram os valores **FRI** marcados em negrito. A coluna **WIL** nas tabelas exibe os resultados das comparações nas quais foi

verificada significância estatística nas diferenças apresentadas, utilizando a seguinte nomenclatura:

- A: Método Co-Training x Self-Training
- B: Método Co-Training x Supervisionado
- C: Método Self-Training x Supervisionado

Nas Tabelas A.25, A.26 e A.27, tem-se os resultados de cada métrica de avaliação para o aprendizado supervisionado nos três níveis da hierarquia e os métodos multirrótulo utilizados: BR, CC e LP. A métrica que obteve o melhor resultado entre os três níveis, teve o seu valor sombreado em cada abordagem multirrótulo. Vale salientar que não se gera resultado para a métrica HiLoss, referente à *Hierarchical Loss*, no primeiro nível, pois, como não há uma hierarquia de classes, a métrica não se é utilizada, pois, este nível é tratado como problema de classificação plana. Nas tabelas deste estudo, a HiLoss, no nível 1, está em destaque. A Tabela A.25 é referente ao aprendizado supervisionado, já a Tabela A.26 é ao semissupervisionado visão simples e a Tabela A.27 pertence ao semissupervisionado multivisão.

Tabela A.1: Média e o desvio padrão dos métodos BR, CC e LP supervisionados Nível 1.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,30026	0,021107	0,30531	0,020231	0,3037	0,022965	0,301	
SAcc↑	0,24901	0,025806	0,3223	0,029527	0,33002	0,042241	0,001	AB
Acc↑	0,40954	0,029186	0,44507	0,035005	0,44479	0,039093	0,002	AB
1-Err↓	0,4384	0,035909	0,44578	0,043338	0,44448	0,044173	0,199	
AvPrec↑	0,71052	0,02382	0,7059	0,024953	0,71522	0,026646	0,670	
RLoss↓	0,33575	0,033202	0,34119	0,028983	0,3084	0,03172	0,169	
CHURCH								
HLoss↓	0,30638	0,005855	0,31114	0,011799	0,31488	0,009733	0,83	
SAcc↑	0,24453	0,020437	0,32408	0,020286	0,31168	0,027211	0,000	AB
Acc↑	0,38425	0,029799	0,43505	0,019715	0,42231	0,019339	0,002	AB
1-Err↓	0,46699	0,026362	0,44938	0,033707	0,45594	0,024281	0,255	
AvPrec↑	0,70038	0,013685	0,70836	0,019548	0,71029	0,013195	0,273	
RLoss↓	0,34635	0,016852	0,3406	0,025212	0,30542	0,011755	0,003	BC
DERISI								
HLoss↓	0,30267	0,020043	0,31367	0,012846	0,31199	0,019275	0,199	
SAcc↑	0,27231	0,034977	0,32489	0,019344	0,31989	0,035748	0,002	AB
Acc↑	0,38703	0,033541	0,4342	0,017122	0,43058	0,033952	0,007	AB
1-Err↓	0,45978	0,041712	0,46145	0,024739	0,45404	0,038023	0,704	
AvPrec↑	0,70525	0,017521	0,70424	0,012817	0,71307	0,023726	0,497	
RLoss↓	0,33754	0,017503	0,34012	0,017747	0,30332	0,028471	0,006	BC
EISEN								
HLoss↓	0,30013	0,021783	0,31047	0,014509	0,31771	0,024305	0,078	
SAcc↑	0,25836	0,0387	0,31918	0,033326	0,30587	0,040165	0,007	AB
Acc↑	0,4272	0,040405	0,44433	0,029226	0,42693	0,039391	0,497	
1-Err↓	0,42617	0,026017	0,44645	0,040021	0,45971	0,036844	0,041	B
AvPrec↑	0,72413	0,017824	0,7141	0,023151	0,7108	0,022766	0,150	
RLoss↓	0,32024	0,023601	0,33163	0,025441	0,31464	0,031188	0,122	
GASCHI								
HLoss↓	0,2871	0,022688	0,28876	0,023344	0,30027	0,017778	0,150	
SAcc↑	0,28715	0,050247	0,36127	0,04817	0,33833	0,023064	0,002	AB
Acc↑	0,43584	0,050357	0,48529	0,03664	0,45527	0,024405	0,020	AC
1-Err↓	0,41648	0,041988	0,40915	0,036024	0,43568	0,031451	0,232	
AvPrec↑	0,7285	0,026811	0,73358	0,022621	0,72122	0,015339	0,202	
RLoss↓	0,31438	0,035279	0,3125	0,030554	0,30317	0,021713	0,905	
GASCH2								
HLoss↓	0,29865	0,013556	0,31104	0,018174	0,30376	0,014544	0,008	A
SAcc↑	0,25991	0,029335	0,3276	0,032716	0,33083	0,024907	0,000	AB
Acc↑	0,39899	0,031733	0,43786	0,035061	0,44447	0,027035	0,000	AB
1-Err↓	0,44535	0,035867	0,44705	0,041412	0,43722	0,030991	0,199	
AvPrec↑	0,70782	0,021759	0,70594	0,024603	0,71957	0,020106	0,061	
RLoss↓	0,34154	0,029263	0,34445	0,028571	0,29989	0,024302	0,001	BC
SPO								
HLoss↓	0,31093	0,0139	0,31905	0,015434	0,30161	0,013655	0,082	
SAcc↑	0,22607	0,026254	0,3051	0,022447	0,33523	0,026542	0,000	ABC
Acc↑	0,39556	0,027183	0,4196	0,023057	0,44945	0,026616	0,007	ABC
1-Err↓	0,45759	0,026094	0,46175	0,028922	0,43608	0,032659	0,266	
AvPrec↑	0,69804	0,019053	0,69435	0,018805	0,72258	0,01903	0,122	
RLoss↓	0,35414	0,028911	0,35816	0,027406	0,29985	0,018526	0,006	BC

Tabela A.2: Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado visão simples Nível 1.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,30496	0,016664	0,31492	0,014715	0,30986	0,012821	0,717	
SAcc↑	0,24821	0,032481	0,30836	0,033566	0,31896	0,02819	0,000	AB
Acc↑	0,39311	0,031082	0,43025	0,029827	0,43749	0,025464	0,061	
1-Err↓	0,46374	0,021278	0,46215	0,030664	0,45596	0,03334	0,575	
AvPrec↑	0,69531	0,012244	0,69614	0,016638	0,70814	0,019134	0,122	
RLoss↓	0,35597	0,022	0,35287	0,019908	0,31771	0,021231	0,014	BC
CHURCH								
HLoss↓	0,30126	0,012949	0,30898	0,010788	0,31528	0,010163	0,027	B
SAcc↑	0,2617	0,045749	0,32736	0,024455	0,31005	0,027237	0,007	ABC
Acc↑	0,37801	0,027238	0,44163	0,018754	0,42203	0,020514	0,003	ABC
1-Err↓	0,46576	0,023299	0,4461	0,029097	0,45717	0,026864	0,125	
AvPrec↑	0,70224	0,012898	0,71115	0,016222	0,70984	0,01507	0,061	
RLoss↓	0,34416	0,016563	0,33536	0,023179	0,30551	0,013715	0,002	BC
DERISI								
HLoss↓	0,30984	0,015908	0,31204	0,019389	0,31414	0,021659	0,741	
SAcc↑	0,24647	0,022832	0,31627	0,033958	0,3121	0,039728	0,002	AB
Acc↑	0,38053	0,008506	0,43208	0,032073	0,42578	0,038404	0,008	AB
1-Err↓	0,47661	0,046032	0,4594	0,036422	0,45609	0,047239	0,202	
AvPrec↑	0,69351	0,022424	0,70433	0,020407	0,7111	0,025776	0,045	B
RLoss↓	0,34866	0,033831	0,33552	0,025428	0,30558	0,025133	0,020	BC
EISEN								
HLoss↓	0,31091	0,021053	0,31743	0,02338	0,31327	0,019712	0,497	
SAcc↑	0,24443	0,053034	0,29513	0,037964	0,31725	0,03077	0,014	AB
Acc↑	0,39995	0,04417	0,43393	0,027698	0,43737	0,032556	0,045	A
1-Err↓	0,45719	0,044445	0,45532	0,049006	0,45336	0,034776	0,905	
AvPrec↑	0,7044	0,012546	0,70165	0,027286	0,71406	0,018322	0,122	
RLoss↓	0,3387	0,016274	0,34507	0,030695	0,31384	0,019489	0,025	BC
GASCHI								
HLoss↓	0,29118	0,016564	0,29468	0,010229	0,30068	0,020343	0,301	
SAcc↑	0,27652	0,031355	0,33963	0,022216	0,33869	0,034321	0,001	AB
Acc↑	0,43472	0,029257	0,47052	0,019819	0,4565	0,030666	0,003	A
1-Err↓	0,43815	0,027181	0,42674	0,024595	0,4357	0,029938	0,529	
AvPrec↑	0,72073	0,013988	0,7213	0,015177	0,72028	0,014314	0,670	
RLoss↓	0,31427	0,020769	0,32213	0,029115	0,30328	0,017027	0,584	
GASCH2								
HLoss↓	0,30446	0,013697	0,31123	0,016068	0,30955	0,010361	0,575	
SAcc↑	0,2452	0,037509	0,32437	0,025721	0,32142	0,018549	0,002	AB
Acc↑	0,38607	0,0402	0,43802	0,02348	0,43565	0,026298	0,008	AB
1-Err↓	0,46577	0,045107	0,45274	0,02689	0,44825	0,033685	0,905	
AvPrec↑	0,69507	0,024186	0,70537	0,015025	0,71264	0,014777	0,202	
RLoss↓	0,35187	0,02563	0,34175	0,020059	0,30684	0,01314	0,001	BC
SPO								
HLoss↓	0,30823	0,012599	0,31314	0,014056	0,30317	0,010671	0,082	
SAcc↑	0,22653	0,029173	0,3142	0,029722	0,32779	0,02273	0,000	AB
Acc↑	0,36807	0,023785	0,42771	0,028405	0,44512	0,02224	0,006	AB
1-Err↓	0,4621	0,026997	0,45761	0,032371	0,43729	0,024878	0,097	
AvPrec↑	0,69717	0,01368	0,69354	0,024252	0,71936	0,01594	0,002	BC
RLoss↓	0,35286	0,015334	0,3594	0,026017	0,30301	0,020588	0,001	BC

Tabela A.3: Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado multivisão Nível 1.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,29863	0,012562	0,29989	0,020076	0,30403	0,017814	0,232	
SAcc↑	0,26539	0,035319	0,29119	0,043969	0,32917	0,027902	0,000	ABC
Acc↑	0,38132	0,038807	0,40235	0,048314	0,44539	0,030619	0,001	BC
1-Err↓	0,42978	0,03154	0,43757	0,038109	0,44452	0,037325	0,784	
AvPrec↑	0,7156	0,01682	0,70815	0,020209	0,71492	0,0209	0,209	
RLoss↓	0,32529	0,022599	0,33223	0,038319	0,31103	0,023918	0,301	
CHURCH								
HLoss↓	0,30053	0,010328	0,30713	0,012387	0,31633	0,00998	0,014	B
SAcc↑	0,25232	0,042536	0,29826	0,033104	0,30841	0,027992	0,003	AB
Acc↑	0,38692	0,024978	0,41337	0,044123	0,41957	0,017543	0,025	B
1-Err↓	0,45838	0,029654	0,45633	0,034592	0,45963	0,022666	0,974	
AvPrec↑	0,70214	0,013801	0,70612	0,018204	0,70763	0,015036	0,301	
RLoss↓	0,34732	0,018681	0,33981	0,021051	0,30824	0,013722	0,000	BC
DERISI								
HLoss↓	0,30182	0,020977	0,30733	0,024839	0,31279	0,018951	0,071	
SAcc↑	0,25591	0,037148	0,29045	0,04837	0,31292	0,034174	0,007	B
Acc↑	0,36868	0,009454	0,40309	0,052379	0,42753	0,032237	0,014	B
1-Err↓	0,46143	0,045193	0,45694	0,049668	0,45403	0,038332	0,794	
AvPrec↑	0,70418	0,022773	0,70875	0,027888	0,71244	0,022248	0,407	
RLoss↓	0,3319	0,022582	0,32175	0,032702	0,30315	0,024479	0,020	BC
EISEN								
HLoss↓	0,30203	0,019496	0,3035	0,023539	0,30097	0,017431	0,741	
SAcc↑	0,25389	0,033928	0,28182	0,044222	0,33381	0,011118	0,002	BC
Acc↑	0,37593	0,037858	0,41245	0,04754	0,45673	0,02723	0,025	BC
1-Err↓	0,40967	0,039358	0,41988	0,042205	0,44143	0,040933	0,273	
AvPrec↑	0,7265	0,019476	0,72042	0,023491	0,71253	0,024361	0,202	
RLoss↓	0,32219	0,032526	0,32228	0,022328	0,31487	0,029822	0,407	
GASCHI								
HLoss↓	0,27962	0,015774	0,27842	0,013478	0,30097	0,017431	0,002	BC
SAcc↑	0,28962	0,030477	0,32897	0,032986	0,33381	0,011118	0,004	AB
Acc↑	0,41158	0,0283	0,44915	0,030823	0,45673	0,02723	0,007	AB
1-Err↓	0,40381	0,023449	0,38747	0,031718	0,44143	0,040933	0,052	
AvPrec↑	0,73447	0,014552	0,74247	0,016812	0,71866	0,020682	0,273	
RLoss↓	0,30335	0,024294	0,29292	0,036077	0,30651	0,011666	0,150	
GASCH2								
HLoss↓	0,29509	0,009942	0,29768	0,014156	0,30619	0,014969	0,019	BC
SAcc↑	0,27498	0,02076	0,30726	0,03351	0,32554	0,031015	0,001	AB
Acc↑	0,37532	0,024532	0,41541	0,040412	0,43943	0,028569	0,007	AB
1-Err↓	0,43233	0,025828	0,43076	0,033271	0,44293	0,032297	0,741	
AvPrec↑	0,71183	0,01266	0,71327	0,020536	0,71436	0,01983	0,670	
RLoss↓	0,32932	0,01408	0,32649	0,022555	0,30872	0,020523	0,301	
SPO								
HLoss↓	0,30326	0,012776	0,30788	0,018876	0,30399	0,014577	0,407	
SAcc↑	0,25421	0,04978	0,2782	0,036028	0,33233	0,029817	0,000	BC
Acc↑	0,37444	0,032482	0,38932	0,038784	0,44621	0,025643	0,002	BC
1-Err↓	0,45222	0,024163	0,46091	0,036153	0,44022	0,022478	0,285	
AvPrec↑	0,70673	0,022437	0,69887	0,02368	0,71957	0,015923	0,273	
RLoss↓	0,3282	0,028791	0,3437	0,030194	0,30463	0,023805	0,082	

Tabela A.4: Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado visão simples Nível 1.

	50		75		
CELLCYCLE					
Medida	Média	Des. Padrão	Média	Des. Padrão	WIL
HLoss↓	0,30496	0,016664	0,31088	0,015345	0,173
SAcc↑	0,31896	0,02819	0,31734	0,028836	0,799
Acc↑	0,43749	0,025464	0,43566	0,02906	0,799
1-Err↓	0,45596	0,03334	0,4572	0,036664	0,799
AvPrec↑	0,70814	0,019134	0,70685	0,02423	0,799
RLoss↓	0,31771	0,021231	0,32076	0,030785	0,858
CHURCH					
HLoss↓	0,30126	0,012949	0,30579	0,012421	0,515
SAcc↑	0,32736	0,024455	0,31629	0,033732	0,097
Acc↑	0,44163	0,018754	0,42628	0,021977	0,083
1-Err↓	0,4461	0,029097	0,46127	0,02006	0,113
AvPrec↑	0,71115	0,016222	0,70636	0,017115	0,508
RLoss↓	0,30551	0,013715	0,30961	0,017623	0,066
DERISI					
HLoss↓	0,30984	0,015908	0,31475	0,017044	0,333
SAcc↑	0,31627	0,033958	0,31003	0,030595	0,445
Acc↑	0,43208	0,032073	0,42404	0,030017	0,508
1-Err↓	0,45609	0,047239	0,45853	0,037064	0,359
AvPrec↑	0,7111	0,025776	0,7089	0,019399	0,575
RLoss↓	0,30558	0,025133	0,30909	0,01963	0,386
EISEN					
HLoss↓	0,31091	0,021053	0,31928	0,03149	0,515
SAcc↑	0,31725	0,03077	0,30839	0,054912	0,398
Acc↑	0,43737	0,032556	0,433	0,051257	0,683
1-Err↓	0,45336	0,034776	0,4578	0,054255	0,674
AvPrec↑	0,71406	0,018322	0,71046	0,035453	0,799
RLoss↓	0,31384	0,019489	0,3188	0,04127	0,721
GASCHI					
HLoss↓	0,29118	0,016564	0,29855	0,015073	0,374
SAcc↑	0,33963	0,022216	0,33092	0,034992	0,646
Acc↑	0,47052	0,019819	0,45939	0,018591	0,203
1-Err↓	0,42674	0,024595	0,43451	0,02012	0,260
AvPrec↑	0,7213	0,015177	0,71846	0,011903	0,721
RLoss↓	0,30328	0,017027	0,30653	0,022143	0,721
GASCH2					
HLoss↓	0,30446	0,013697	0,30956	0,012995	0,333
SAcc↑	0,32437	0,025721	0,3198	0,024321	0,575
Acc↑	0,43802	0,02348	0,43565	0,026298	0,444
1-Err↓	0,44905	0,028881	0,44825	0,033685	0,608
AvPrec↑	0,71264	0,014777	0,71195	0,018612	0,878
RLoss↓	0,30684	0,01314	0,31212	0,021143	0,508
SPO					
HLoss↓	0,30317	0,010671	0,31061	0,017416	0,109
SAcc↑	0,32779	0,02273	0,32034	0,032558	0,409
Acc↑	0,44512	0,02224	0,43496	0,030593	0,114
1-Err↓	0,43729	0,024878	0,44971	0,034019	0,139
AvPrec↑	0,71936	0,01594	0,71396	0,021751	0,169
RLoss↓	0,30301	0,020588	0,3101	0,026809	0,445

Tabela A.5: Média e o desvio padrão dos percentuais iniciais de exemplos rotulados para o método semissupervisionado multivisão Nível 1.

	50		75		
CELLCYCLE					
Medida	Média	Des. Padrão	Média	Des. Padrão	WIL
HLoss↓	0,29863	0,012562	0,30427	0,015198	0,074
SAcc↑	0,32917	0,027902	0,31527	0,023615	0,114
Acc↑	0,44539	0,030619	0,43366	0,02634	0,386
1-Err↓	0,42978	0,03154	0,44657	0,031047	0,092
AvPrec↑	0,7156	0,01682	0,70767	0,022295	0,333
RLoss↓	0,31103	0,023918	0,31816	0,025079	0,333
CHURCH					
HLoss↓	0,30053	0,010328	0,30137	0,010412	0,767
SAcc↑	0,30841	0,027992	0,30349	0,031231	0,089
Acc↑	0,41957	0,017543	0,41575	0,022855	0,152
1-Err↓	0,45633	0,034592	0,45838	0,029654	0,906
AvPrec↑	0,70763	0,015036	0,70466	0,015813	0,059
RLoss↓	0,30824	0,013722	0,3125	0,016791	0,038
DERISI					
HLoss↓	0,30182	0,020977	0,3079	0,021354	0,005
SAcc↑	0,31292	0,034174	0,31168	0,029023	0,594
Acc↑	0,42753	0,032237	0,42705	0,030887	0,878
1-Err↓	0,45403	0,038332	0,45611	0,036135	0,859
AvPrec↑	0,71244	0,022248	0,70937	0,020643	0,575
RLoss↓	0,30315	0,024479	0,3113	0,025377	0,126
EISEN					
HLoss↓	0,30097	0,017431	0,31347	0,018637	0,169
SAcc↑	0,33381	0,011118	0,29067	0,035633	0,009
Acc↑	0,45673	0,02723	0,41402	0,034766	0,007
1-Err↓	0,40967	0,039358	0,4642	0,035908	0,017
AvPrec↑	0,7265	0,019476	0,70109	0,024154	0,059
RLoss↓	0,31487	0,029822	0,32606	0,027697	0,374
GASCHI					
HLoss↓	0,27842	0,013478	0,291	0,015395	0,013
SAcc↑	0,33381	0,011118	0,3248	0,018104	0,262
Acc↑	0,45673	0,02723	0,44656	0,017418	0,169
1-Err↓	0,38747	0,031718	0,41279	0,028408	0,038
AvPrec↑	0,74247	0,016812	0,72428	0,01987	0,017
RLoss↓	0,29292	0,036077	0,30651	0,011666	0,169
GASCH2					
HLoss↓	0,29509	0,009942	0,30306	0,014325	0,021
SAcc↑	0,32554	0,031015	0,32023	0,023509	0,575
Acc↑	0,43943	0,028569	0,43442	0,023173	0,445
1-Err↓	0,43076	0,033271	0,44495	0,026715	0,114
AvPrec↑	0,71436	0,01983	0,71182	0,017076	0,646
RLoss↓	0,30872	0,020523	0,3137	0,022103	0,241
SPO					
HLoss↓	0,30326	0,012776	0,30513	0,013395	0,798
SAcc↑	0,33233	0,029817	0,31827	0,032959	0,126
Acc↑	0,44621	0,025643	0,43213	0,031359	0,169
1-Err↓	0,44022	0,022478	0,4526	0,020489	0,358
AvPrec↑	0,71957	0,015923	0,71062	0,0218	0,333
RLoss↓	0,30463	0,023805	0,3144	0,021576	0,203

Tabela A.6: Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado visão simples Nível 1.

	20	30	40	50		
CELLCYCLE						
Medida	Média	Média	Média	Média	FRI	WIL
HLoss↓	0,30825	0,31527	0,30496	0,31088	0,285	
SAcc↑	0,31734	0,31325	0,31896	0,3128	0,975	
Acc↑	0,43519	0,42627	0,43749	0,43566	0,564	
1-Err↓	0,4633	0,46451	0,45596	0,4572	0,965	
AvPrec↑	0,70148	0,6996	0,70814	0,70685	0,462	
RLoss↓	0,32862	0,33332	0,31771	0,32076	0,031	BD
CHURCH						
HLoss↓	0,30617	0,30474	0,30547	0,30126	0,194	
SAcc↑	0,32614	0,32736	0,32736	0,31916	0,542	
Acc↑	0,43831	0,43839	0,44163	0,43093	0,801	
1-Err↓	0,44856	0,45225	0,4461	0,45676	0,935	
AvPrec↑	0,70984	0,70777	0,71115	0,70717	0,811	
RLoss↓	0,30551	0,31164	0,30863	0,3088	0,330	
DERISI						
HLoss↓	0,31435	0,30984	0,3186	0,31474	0,860	
SAcc↑	0,30799	0,31627	0,31297	0,31086	0,668	
Acc↑	0,42362	0,43208	0,42483	0,42495	0,782	
1-Err↓	0,45609	0,45772	0,45978	0,45853	0,675	
AvPrec↑	0,7111	0,71073	0,70757	0,70981	0,519	
RLoss↓	0,30558	0,30633	0,31149	0,30625	0,323	
EISEN						
HLoss↓	0,31201	0,31091	0,31662	0,31659	0,671	
SAcc↑	0,30968	0,31725	0,3084	0,30839	0,996	
Acc↑	0,43077	0,43737	0,43026	0,43393	0,724	
1-Err↓	0,45524	0,45336	0,45532	0,45653	0,976	
AvPrec↑	0,71095	0,71406	0,71071	0,71046	0,923	
RLoss↓	0,32051	0,31384	0,31411	0,31694	0,965	
GASCH1						
HLoss↓	0,29454	0,29118	0,29373	0,29373	0,872	
SAcc↑	0,33838	0,33963	0,33869	0,33839	0,896	
Acc↑	0,46766	0,47052	0,46388	0,46671	0,836	
1-Err↓	0,42674	0,42838	0,43451	0,43368	0,787	
AvPrec↑	0,7213	0,71265	0,71945	0,72028	0,916	
RLoss↓	0,30577	0,31323	0,30653	0,30328	0,421	
GASCH2						
HLoss↓	0,30752	0,30823	0,30986	0,30446	0,365	
SAcc↑	0,32359	0,31739	0,32104	0,32437	0,804	
Acc↑	0,43454	0,43077	0,43565	0,43802	0,791	
1-Err↓	0,44945	0,4527	0,44825	0,44944	0,791	
AvPrec↑	0,70998	0,70886	0,71195	0,71264	0,668	
RLoss↓	0,31654	0,31484	0,31148	0,30684	0,197	
SPO						
HLoss↓	0,30472	0,30874	0,30833	0,30317	0,422	
SAcc↑	0,32777	0,32239	0,32777	0,32779	0,958	
Acc↑	0,44319	0,43715	0,43909	0,44512	0,787	
1-Err↓	0,4406	0,44767	0,4464	0,43729	0,679	
AvPrec↑	0,71883	0,71457	0,71538	0,71936	0,816	
RLoss↓	0,30301	0,30825	0,30896	0,30331	0,615	

Tabela A.7: Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado multivisão Nível 1.

	20	30	40	50		
CELLCYCLE						
Medida	Média	Média	Média	Média	FRI	WIL
HLoss↓	0,30546	0,30153	0,29989	0,29863	0,494	
SAcc↑	0,31853	0,32917	0,32795	0,32757	0,641	
Acc↑	0,4347	0,44327	0,44539	0,44449	0,246	
1-Err↓	0,44821	0,44289	0,43757	0,42978	0,596	
AvPrec↑	0,70532	0,71199	0,71492	0,7156	0,095	
RLoss↓	0,32784	0,31716	0,31103	0,31413	0,421	
CHURCH						
HLoss↓	0,30249	0,30073	0,30053	0,30249	0,884	
SAcc↑	0,30553	0,30841	0,3039	0,30841	0,180	
Acc↑	0,41751	0,41928	0,41661	0,41957	0,454	
1-Err↓	0,46168	0,45633	0,45633	0,45757	0,473	
AvPrec↑	0,70631	0,70763	0,70612	0,70749	0,989	
RLoss↓	0,31011	0,30824	0,31145	0,30911	0,421	
DERISI						
HLoss↓	0,30182	0,3079	0,30733	0,3087	0,078	
SAcc↑	0,31292	0,31292	0,31169	0,30758	0,907	
Acc↑	0,42753	0,42705	0,42595	0,42365	0,993	
1-Err↓	0,45486	0,45611	0,45403	0,45483	0,975	
AvPrec↑	0,7121	0,71107	0,71244	0,71154	0,948	
RLoss↓	0,30396	0,30491	0,30315	0,30459	0,971	
EISEN						
HLoss↓	0,30346	0,3074	0,30203	0,30488	0,451	
SAcc↑	0,30647	0,297	0,31219	0,31473	0,752	
Acc↑	0,42954	0,42535	0,43446	0,43295	0,941	
1-Err↓	0,40967	0,44197	0,42549	0,43889	0,038	C
AvPrec↑	0,7265	0,71346	0,72	0,71543	0,293	
RLoss↓	0,31616	0,3254	0,31487	0,32228	0,564	
GASCH1						
HLoss↓	0,27977	0,28845	0,28226	0,27842	0,323	
SAcc↑	0,32928	0,33176	0,33177	0,33381	0,959	
Acc↑	0,44915	0,45673	0,44966	0,44803	0,753	
1-Err↓	0,38747	0,40627	0,39975	0,39153	0,754	
AvPrec↑	0,74247	0,73031	0,73709	0,74042	0,401	
RLoss↓	0,29514	0,30651	0,29292	0,29909	0,145	
GASCH2						
HLoss↓	0,29895	0,30166	0,29598	0,29509	0,147	
SAcc↑	0,32022	0,32145	0,32554	0,32023	0,854	
Acc↑	0,43181	0,43304	0,43943	0,43442	0,971	
1-Err↓	0,44704	0,43645	0,43279	0,43076	0,258	
AvPrec↑	0,71039	0,71006	0,71436	0,71327	0,948	
RLoss↓	0,31576	0,31756	0,30872	0,3137	0,941	
SPO						
HLoss↓	0,30326	0,30494	0,30441	0,30399	0,737	
SAcc↑	0,33232	0,31909	0,32365	0,33233	0,355	
Acc↑	0,44491	0,43196	0,44088	0,44621	0,644	
1-Err↓	0,44434	0,45222	0,44267	0,44022	0,566	
AvPrec↑	0,71676	0,71062	0,71658	0,71957	0,874	
RLoss↓	0,3089	0,31337	0,30704	0,30463	0,860	

Tabela A.8: Média e desvio padrão das abordagens semissupervisionada multivisão *versus* semissupervisionada visão simples *versus* abordagem supervisionada Nível 1.

	Co-Training		Self-Training		Supervisionado			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,30403	0,017814	0,30986	0,012821	0,3037	0,022965	0,273	
SAcc↑	0,32917	0,027902	0,31896	0,02819	0,33002	0,042241	0,368	
Acc↑	0,44539	0,030619	0,43749	0,025464	0,44479	0,039093	0,452	
1-Err↓	0,44452	0,037325	0,45596	0,03334	0,44448	0,044173	0,285	
AvPrec↑	0,71492	0,0209	0,70814	0,019134	0,71522	0,026646	0,122	
RLoss↓	0,31103	0,023918	0,31771	0,021231	0,3084	0,03172	0,122	
CHURCH								
HLoss↓	0,31633	0,00998	0,31528	0,010163	0,31488	0,009733	0,125	
SAcc↑	0,30841	0,027992	0,31005	0,027237	0,31168	0,027211	0,214	
Acc↑	0,41957	0,017543	0,42203	0,020514	0,42231	0,019339	0,192	
1-Err↓	0,45963	0,022666	0,45717	0,026864	0,45594	0,024281	0,087	
AvPrec↑	0,70763	0,015036	0,70984	0,01507	0,71029	0,013195	0,087	
RLoss↓	0,30824	0,013722	0,30551	0,013715	0,30542	0,011755	0,128	
DERISI								
HLoss↓	0,31279	0,018951	0,31414	0,021659	0,31199	0,019275	0,186	
SAcc↑	0,31292	0,034174	0,3121	0,039728	0,31989	0,035748	0,184	
Acc↑	0,42753	0,032237	0,42578	0,038404	0,43058	0,033952	0,452	
1-Err↓	0,45403	0,038332	0,45609	0,047239	0,45404	0,038023	0,755	
AvPrec↑	0,71244	0,022248	0,7111	0,025776	0,71307	0,023726	0,926	
RLoss↓	0,30315	0,024479	0,30558	0,025133	0,30332	0,028471	0,497	
EISEN								
HLoss↓	0,30097	0,017431	0,31327	0,019712	0,31771	0,024305	0,497	
SAcc↑	0,33381	0,011118	0,31725	0,03077	0,30587	0,040165	0,033	B
Acc↑	0,45673	0,02723	0,43737	0,032556	0,42693	0,039391	0,150	
1-Err↓	0,44143	0,040933	0,45336	0,034776	0,45971	0,036844	0,656	
AvPrec↑	0,71253	0,024361	0,71406	0,018322	0,7108	0,022766	0,905	
RLoss↓	0,31487	0,029822	0,31384	0,019489	0,31464	0,031188	0,905	
GASCH1								
HLoss↓	0,30097	0,017431	0,30068	0,020343	0,30027	0,017778	0,836	
SAcc↑	0,33381	0,011118	0,33869	0,034321	0,33833	0,023064	0,922	
Acc↑	0,45673	0,02723	0,4565	0,030666	0,45527	0,024405	0,670	
1-Err↓	0,44143	0,040933	0,4357	0,029938	0,43568	0,031451	0,975	
AvPrec↑	0,71866	0,020682	0,72028	0,014314	0,72122	0,015339	0,741	
RLoss↓	0,30651	0,011666	0,30328	0,017027	0,30317	0,021713	0,614	
GASCH2								
HLoss↓	0,30619	0,014969	0,30955	0,010361	0,30376	0,014544	0,122	
SAcc↑	0,32554	0,031015	0,32142	0,018549	0,33083	0,024907	0,469	
Acc↑	0,43943	0,028569	0,43565	0,026298	0,44447	0,027035	0,452	
1-Err↓	0,44293	0,032297	0,44825	0,033685	0,43722	0,030991	0,452	
AvPrec↑	0,71436	0,01983	0,71264	0,014777	0,71957	0,020106	0,150	
RLoss↓	0,30872	0,020523	0,30684	0,01314	0,29989	0,024302	0,061	
SPO								
HLoss↓	0,30399	0,014577	0,30317	0,010671	0,30161	0,013655	0,209	
SAcc↑	0,33233	0,029817	0,32779	0,02273	0,33523	0,026542	0,349	
Acc↑	0,44621	0,025643	0,44512	0,02224	0,44945	0,026616	0,614	
1-Err↓	0,44022	0,022478	0,43729	0,024878	0,43608	0,032659	0,273	
AvPrec↑	0,71957	0,015923	0,71936	0,01594	0,72258	0,01903	0,497	
RLoss↓	0,30463	0,023805	0,30301	0,020588	0,29985	0,018526	0,452	

Tabela A.9: Média e o desvio padrão dos métodos BR, CC e LP supervisionados Nível 2.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,11943	0,002891	0,12989	0,003205	0,13385	0,004749	0,000	ABC
SAcc↑	0,06413	0,019342	0,09205	0,010447	0,06741	0,016791	0,002	AC
Acc↑	0,2377	0,021431	0,31249	0,01634	0,23521	0,018367	0,002	AC
1-Err↓	0,52879	0,032814	0,52884	0,032849	0,58481	0,03326	0,000	ABC
AvPrec↑	0,38953	0,017381	0,39346	0,01991	0,37742	0,018147	0,003	BC
RLoss↓	0,44646	0,019025	0,44342	0,010526	0,46386	0,016027	0,001	BC
HiLoss↓	1,93689	0,061442	2,06468	0,048922	2,13534	0,060537	0,000	ABC
CHURCH								
HLoss↓	0,11509	0,005026	0,13171	0,00322	0,13608	0,003519	0,000	ABC
SAcc↑	0,02941	0,012489	0,06803	0,017845	0,07039	0,019882	0,001	AB
Acc↑	0,19952	0,014033	0,014131	0,014131	0,22568	0,009591	0,000	ABC
1-Err↓	0,61525	0,047886	0,49243	0,059386	0,73737	0,025246	0,000	ABC
AvPrec↑	0,37755	0,014932	0,3801	0,015142	0,27702	0,016314	0,001	BC
RLoss↓	0,39963	0,019711	0,4041	0,017415	0,41615	0,016472	0,007	ABC
HiLoss↓	1,77092	0,091093	2,07373	0,063163	2,16219	0,036303	0,000	ABC
DERISI								
HLoss↓	0,11921	0,004568	0,13543	0,004167	0,13608	0,003329	0,000	ABC
SAcc↑	0,05284	0,012591	0,014553	0,014553	0,05284	0,012591	0,002	AC
Acc↑	0,24509	0,006153	0,24489	0,007627	0,21856	0,009332	0,001	BC
1-Err↓	0,54996	0,023636	0,55006	0,023636	0,60249	0,03284	0,000	ABC
AvPrec↑	0,40288	0,005262	0,40198	0,005793	0,36261	0,017923	0,000	BC
RLoss↓	0,36608	0,006928	0,36737	0,010536	0,44479	0,015129	0,000	BC
HiLoss↓	1,85887	0,076045	2,71693	0,078922	3,4836	0,140959	0,000	ABC
EISEN								
HLoss↓	0,12553	0,00525	0,13329	0,008336	0,13631	0,005309	0,006	AB
SAcc↑	0,0601	0,021745	0,08106	0,025438	0,06138	0,020248	0,061	
Acc↑	0,24065	0,027755	0,25923	0,017776	0,24447	0,018655	0,150	
1-Err↓	0,54209	0,053618	0,54212	0,053591	0,59338	0,044509	0,000	BC
AvPrec↑	0,38005	0,027019	0,38011	0,026989	0,36936	0,024987	0,002	ABC
RLoss↓	0,44289	0,026012	0,44808	0,015885	0,45627	0,01681	0,082	
HiLoss↓	1,9835	0,059084	2,09308	0,114469	2,18685	0,095592	0,000	ABC
GASCHI								
HLoss↓	0,11952	0,004222	0,12461	0,004985	0,13004	0,004229	0,001	ABC
SAcc↑	0,05064	0,019084	0,08879	0,019373	0,07972	0,014084	0,001	AB
Acc↑	0,27168	0,020832	0,28311	0,020603	0,25531	0,018133	0,061	
1-Err↓	0,5425	0,031789	0,50823	0,045094	0,62802	0,049452	0,001	BC
AvPrec↑	0,40687	0,017909	0,412	0,016673	0,35792	0,022652	0,001	BC
RLoss↓	0,41875	0,016838	0,43361	0,010368	0,45408	0,019396	0,006	BC
HiLoss↓	1,95413	0,06125	2,03597	0,046609	2,1018	0,062314	0,001	ABC
GASCH2								
HLoss↓	0,11862	0,004348	0,13091	0,006069	0,13229	0,003229	0,001	AB
SAcc↑	0,06839	0,013388	0,08636	0,02299	0,07167	0,015176	0,378	
Acc↑	0,23366	0,016746	0,24908	0,027871	0,23552	0,014537	0,670	
1-Err↓	0,55412	0,024786	0,5497	0,084297	0,60712	0,025867	0,006	B
AvPrec↑	0,38262	0,012604	0,38264	0,012598	0,37072	0,013181	0,000	BC
RLoss↓	0,45665	0,016654	0,45857	0,018071	0,45833	0,013734	0,905	
HiLoss↓	1,90211	0,06405	2,04241	0,083617	2,10998	0,044395	0,000	ABC
SPO								
HLoss↓	0,11951	0,004361	0,13288	0,0053	0,13215	0,003158	0,000	AB
SAcc↑	0,06941	0,025435	0,06947	0,02541	0,06526	0,014656	0,307	
Acc↑	0,23755	0,015414	0,24698	0,020377	0,23727	0,01337	0,273	
1-Err↓	0,53528	0,027494	0,53534	0,027479	0,58611	0,033642	0,000	ABC
AvPrec↑	0,38706	0,016968	0,3871	0,016956	0,37511	0,019658	0,002	BC
RLoss↓	0,39496	0,018052	0,39438	0,018015	0,45543	0,016279	0,000	BC
HiLoss↓	1,91851	0,066359	2,05205	0,084504	2,12233	0,055361	0,000	ABC

Tabela A.10: Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado visão simples Nível 2.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,12089	0,003523	0,12732	0,005322	0,13504	0,003732	0,000	ABC
SAcc↑	0,03469	0,01352	0,06955	0,018537	0,06373	0,017756	0,000	AB
Acc↑	0,23239	0,015996	0,24043	0,017629	0,22939	0,021484	0,670	
1-Err↓	0,64538	0,028237	0,63923	0,066227	0,78724	0,015081	0,001	BC
AvPrec↑	0,36097	0,01484	0,35333	0,019747	0,32935	0,016	0,008	BC
RLoss↓	0,44834	0,022058	0,44667	0,019982	0,43661	0,033221	0,741	
HiLoss↓	1,94968	0,063299	2,04582	0,053768	2,146	0,044964	0,000	ABC
CHURCH								
HLoss↓	0,11519	0,003709	0,13175	0,003565	0,13606	0,004024	0,000	AB
SAcc↑	0,02695	0,009696	0,07828	0,015646	0,07203	0,008681	0,000	AB
Acc↑	0,21945	0,019917	0,2416	0,015599	0,22791	0,010218	0,007	AC
1-Err↓	0,63249	0,049041	0,73126	0,073419	0,78075	0,023825	0,001	AB
AvPrec↑	0,33088	0,018426	0,34474	0,016844	0,32774	0,010105	0,045	B
RLoss↓	0,43799	0,027888	0,44076	0,031989	0,43856	0,03976	0,497	
HiLoss↓	1,79883	0,084266	2,05531	0,067732	2,16833	0,038796	0,000	ABC
DERISI								
HLoss↓	0,12037	0,006614	0,12687	0,004929	0,13615	0,002347	0,000	ABC
SAcc↑	0,0278	0,016809	0,07139	0,013124	0,05162	0,011721	0,001	ABC
Acc↑	0,23536	0,012501	0,23631	0,014108	0,22521	0,012364	0,045	B
1-Err↓	0,64149	0,04362	0,61937	0,040607	0,80716	0,02735	0,000	BC
AvPrec↑	0,33363	0,020902	0,34285	0,024147	0,31705	0,017099	0,045	C
RLoss↓	0,43161	0,018101	0,43409	0,015561	0,43199	0,026991	0,905	
HiLoss↓	1,78537	0,134558	2,03733	0,064311	2,17303	0,052152	0,000	ABC
EISEN								
HLoss↓	0,1261	0,004498	0,12885	0,008703	0,13745	0,004833	0,045	BC
SAcc↑	0,03731	0,012153	0,07155	0,014255	0,06579	0,016067	0,004	AB
Acc↑	0,2415	0,018764	0,25531	0,020637	0,24253	0,017446	0,670	
1-Err↓	0,65924	0,042026	0,65362	0,076064	0,77261	0,037154	0,002	BC
AvPrec↑	0,3689	0,018984	0,36932	0,027607	0,34454	0,024217	0,025	BC
RLoss↓	0,4294	0,02643	0,42785	0,016676	0,44037	0,02141	0,067	
HiLoss↓	1,98287	0,098088	2,04688	0,073425	2,19197	0,085395	0,002	BC
GASCH1								
HLoss↓	0,12111	0,00426	0,1242	0,003479	0,13004	0,003984	0,007	BC
SAcc↑	0,04371	0,011247	0,08634	0,011993	0,08423	0,015787	0,001	AB
Acc↑	0,25573	0,027178	0,27965	0,014794	0,25693	0,017101	0,150	
1-Err↓	0,63824	0,033196	0,63955	0,054531	0,76263	0,020279	0,001	BC
AvPrec↑	0,37463	0,022312	0,38597	0,012237	0,35475	0,015956	0,007	BC
RLoss↓	0,42842	0,01316	0,42701	0,010808	0,41006	0,029956	0,122	
HiLoss↓	1,95623	0,074494	2,00164	0,035515	2,09853	0,067254	0,006	BC
GASCH2								
HLoss↓	0,12226	0,004504	0,12713	0,006023	0,13399	0,004003	0,000	ABC
SAcc↑	0,03501	0,009261	0,07337	0,018401	0,06598	0,016003	0,001	AB
Acc↑	0,22215	0,018278	0,24715	0,02001	0,23304	0,019455	0,007	A
1-Err↓	0,64787	0,02546	0,63529	0,060432	0,7848	0,02428	0,000	BC
AvPrec↑	0,34714	0,009403	0,35068	0,016669	0,33054	0,020322	0,150	
RLoss↓	0,4516	0,019363	0,44151	0,01999	0,43332	0,043818	0,407	
HiLoss↓	1,94902	0,075801	2,04359	0,057516	2,12425	0,039173	0,003	ABC
SPO								
HLoss↓	0,11927	0,004633	0,12641	0,004739	0,13296	0,005454	0,001	ABC
SAcc↑	0,03259	0,018001	0,07978	0,01325	0,07188	0,018484	0,001	AB
Acc↑	0,2327	0,018783	0,24536	0,018582	0,23948	0,017327	0,273	
1-Err↓	0,65103	0,050768	0,63793	0,063479	0,77962	0,021289	0,000	BC
AvPrec↑	0,34786	0,014436	0,34861	0,023284	0,33826	0,014274	0,122	
RLoss↓	0,44554	0,022152	0,44569	0,020579	0,41965	0,018833	0,008	BC
HiLoss↓	1,85817	0,084422	2,01992	0,080197	2,11778	0,06901	0,001	ABC

Tabela A.11: Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado multivisão Nível 2.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,11143	0,003686	0,1135	0,004988	0,13539	0,004021	0,000	BC
SAcc↑	0,02527	0,009004	0,04255	0,008025	0,06661	0,007974	0,000	ABC
Acc↑	0,21525	0,016962	0,22865	0,014005	0,23057	0,018283	0,202	
1-Err↓	0,56346	0,040344	0,5513	0,03659	0,78724	0,013419	0,000	BC
AvPrec↑	0,34152	0,02285	0,35208	0,02044	0,33184	0,00951	0,407	
RLoss↓	0,45398	0,021027	0,44815	0,016829	0,45357	0,013083	0,332	
HiLoss↓	1,76271	0,099655	1,86755	0,103502	2,13984	0,037392	0,000	BC
CHURCH								
HLoss↓	0,11203	0,00319	0,119	0,007231	0,1361	0,004538	0,000	ABC
SAcc↑	0,02327	0,01077	0,05287	0,028316	0,07778	0,018753	0,000	ABC
Acc↑	0,22073	0,021561	0,22981	0,030762	0,22975	0,021863	0,670	
1-Err↓	0,6046	0,054059	0,58175	0,059112	0,77624	0,024042	0,000	BC
AvPrec↑	0,31446	0,0249	0,35461	0,025276	0,33021	0,022942	0,003	AC
RLoss↓	0,43594	0,028378	0,43757	0,02329	0,43609	0,051903	0,407	
HiLoss↓	1,71399	0,049734	1,85373	0,085071	2,16176	0,044059	0,000	ABC
DERISI								
HLoss↓	0,11343	0,004653	0,11567	0,004972	0,13661	0,003466	0,000	BC
SAcc↑	0,02412	0,009752	0,04102	0,022853	0,05121	0,016177	0,122	
Acc↑	0,23746	0,011478	0,23293	0,017749	0,21464	0,015528	0,007	B
1-Err↓	0,6021	0,04571	0,53571	0,030371	0,80799	0,029136	0,000	ABC
AvPrec↑	0,33147	0,013039	0,34534	0,011179	0,31566	0,022623	0,027	AC
RLoss↓	0,44276	0,012367	0,44451	0,01656	0,44581	0,021293	0,741	
HiLoss↓	1,78757	0,087188	1,90611	0,068172	2,18161	0,063085	0,000	ABC
EISEN								
HLoss↓	0,11636	0,004	0,1191	0,004702	0,13957	0,00459	0,000	BC
SAcc↑	0,03035	0,017845	0,04431	0,019792	0,06646	0,023619	0,038	B
Acc↑	0,21963	0,022465	0,23568	0,019964	0,24469	0,0267	0,012	AB
1-Err↓	0,63516	0,043432	0,59276	0,055831	0,77073	0,038966	0,000	BC
AvPrec↑	0,33873	0,020513	0,35502	0,027071	0,34701	0,024216	0,273	
RLoss↓	0,4429	0,02531	0,43272	0,028273	0,44164	0,021361	0,741	
HiLoss↓	1,8308	0,124863	1,90362	0,132683	2,20079	0,104017	0,000	BC
GASCH1								
HLoss↓	0,11015	0,003087	0,11004	0,003345	0,1333	0,004174	0,001	BC
SAcc↑	0,03102	0,012633	0,04094	0,014461	0,07892	0,015097	0,000	BC
Acc↑	0,23	0,01545	0,23417	0,019438	0,25386	0,020121	0,007	BC
1-Err↓	0,56131	0,045558	0,54665	0,043376	0,76833	0,028826	0,001	BC
AvPrec↑	0,36547	0,019255	0,36747	0,015976	0,35031	0,015286	0,067	
RLoss↓	0,43371	0,016405	0,43547	0,013893	0,43623	0,01308	0,905	
HiLoss↓	1,77775	0,131985	1,78802	0,052924	2,12799	0,04943	0,001	BC
GASCH2								
HLoss↓	0,11128	0,003392	0,11321	0,004238	0,1342	0,006206	0,000	ABC
SAcc↑	0,03174	0,010842	0,05053	0,015414	0,06923	0,01599	0,000	ABC
Acc↑	0,20375	0,010679	0,23229	0,021718	0,23339	0,015425	0,002	AB
1-Err↓	0,55542	0,036202	0,52893	0,046224	0,78559	0,035022	0,000	BC
AvPrec↑	0,33351	0,010058	0,34668	0,019542	0,3324	0,013946	0,150	
RLoss↓	0,46456	0,015385	0,45423	0,024559	0,45391	0,012497	0,273	
HiLoss↓	1,75958	0,079107	1,85134	0,126037	2,12424	0,079315	0,000	BC
SPO								
HLoss↓	0,11189	0,002729	0,11539	0,004652	0,13418	0,005049	0,000	ABC
SAcc↑	0,02845	0,012074	0,04341	0,022987	0,06154	0,015507	0,001	ABC
Acc↑	0,22149	0,015406	0,22463	0,014847	0,23195	0,019524	0,670	
1-Err↓	0,57085	0,049997	0,56885	0,048809	0,78828	0,022623	0,001	BC
AvPrec↑	0,33455	0,022086	0,34453	0,029979	0,33115	0,017273	0,273	
RLoss↓	0,45207	0,026828	0,45493	0,021047	0,44167	0,032209	0,202	
HiLoss↓	1,77835	0,067785	1,86522	0,075442	2,14296	0,04072	0,000	BC

Tabela A.12: Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado visão simples Nível 2.

	50		75		
CELLCYCLE					
Medida	Média	Des. Padrão	Média	Des. Padrão	WIL
HLoss↓	0,12089	0,003523	0,12732	0,005322	0,009
SAcc↑	0,06955	0,018537	0,05924	0,011261	0,093
Acc↑	0,24043	0,017629	0,23476	0,022556	0,799
1-Err↓	0,64538	0,028237	0,63923	0,066227	0,799
AvPrec↑	0,36097	0,01484	0,35333	0,019747	0,646
RLoss↓	0,44667	0,019982	0,43661	0,033221	0,285
HiLoss↓	1,94968	0,063299	2,01061	0,118863	0,059
CHURCH					
HLoss↓	0,11519	0,003709	0,11704	0,005754	0,444
SAcc↑	0,07828	0,015646	0,0754	0,016965	1,000
Acc↑	0,2416	0,015599	0,23631	0,009827	0,333
1-Err↓	0,64108	0,048978	0,63249	0,049041	0,635
AvPrec↑	0,34474	0,016844	0,3412	0,020589	0,959
RLoss↓	0,43856	0,03976	0,43799	0,027888	0,959
HiLoss↓	1,79883	0,084266	1,81968	0,089301	0,333
DERISI					
HLoss↓	0,12037	0,006614	0,12072	0,0044	0,799
SAcc↑	0,07139	0,013124	0,06483	0,015814	0,214
Acc↑	0,23631	0,014108	0,2358	0,014345	0,721
1-Err↓	0,64149	0,04362	0,61937	0,040607	0,374
AvPrec↑	0,33485	0,016376	0,34285	0,024147	0,386
RLoss↓	0,43729	0,012013	0,43161	0,018101	0,359
HiLoss↓	1,81826	0,098545	1,78537	0,134558	0,386
EISEN					
HLoss↓	0,1261	0,004498	0,13301	0,00325	0,005
SAcc↑	0,07155	0,014255	0,06074	0,021328	0,241
Acc↑	0,25531	0,020637	0,2434	0,018533	0,203
1-Err↓	0,65362	0,076064	0,66557	0,041202	0,575
AvPrec↑	0,36932	0,027607	0,36218	0,014077	0,575
RLoss↓	0,42785	0,016676	0,43502	0,036645	0,721
HiLoss↓	1,98287	0,098088	2,08869	0,055902	0,013
GASCH1					
HLoss↓	0,12111	0,00426	0,12448	0,00565	0,139
SAcc↑	0,08634	0,011993	0,074	0,021547	0,203
Acc↑	0,27965	0,014794	0,26283	0,01784	0,093
1-Err↓	0,63824	0,033196	0,63955	0,054531	0,959
AvPrec↑	0,38597	0,012237	0,37402	0,01571	0,114
RLoss↓	0,42701	0,010808	0,41006	0,029956	0,169
HiLoss↓	1,95623	0,074494	1,98319	0,062201	0,203
GASCH2					
HLoss↓	0,12226	0,004504	0,12288	0,005974	0,721
SAcc↑	0,07337	0,018401	0,06033	0,017947	0,074
Acc↑	0,24715	0,02001	0,23997	0,021563	0,721
1-Err↓	0,64787	0,02546	0,63529	0,060432	0,508
AvPrec↑	0,34714	0,009403	0,35068	0,016669	0,445
RLoss↓	0,44151	0,01999	0,43332	0,043818	0,445
HiLoss↓	1,95915	0,06328	1,94902	0,075801	0,799
SPO					
HLoss↓	0,11927	0,004633	0,12314	0,004398	0,047
SAcc↑	0,07978	0,01325	0,06567	0,013145	0,059
Acc↑	0,24536	0,018582	0,2368	0,019454	0,037
1-Err↓	0,65103	0,050768	0,63793	0,063479	0,241
AvPrec↑	0,34853	0,018221	0,34861	0,023284	0,721
RLoss↓	0,41965	0,018833	0,4213	0,020316	0,799
HiLoss↓	1,85817	0,084422	1,94663	0,09741	0,047

Tabela A.13: Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado multivisão Nível 2.

	50		75		
CELLCYCLE					
Medida	Média	Des. Padrão	Média	Des. Padrão	WIL
HLoss↓	0,11143	0,003686	0,11508	0,003396	0,011
SAcc↑	0,06661	0,007974	0,06087	0,010968	0,123
Acc↑	0,23057	0,018283	0,2274	0,016676	0,508
1-Err↓	0,5513	0,03659	0,58408	0,058065	0,139
AvPrec↑	0,35208	0,02044	0,3439	0,01954	0,139
RLoss↓	0,44815	0,016829	0,45841	0,009944	0,203
HiLoss↓	1,83797	0,092142	1,76271	0,099655	0,333
CHURCH					
HLoss↓	0,11203	0,00319	0,11459	0,003832	0,017
SAcc↑	0,07778	0,018753	0,07777	0,020532	0,889
Acc↑	0,22975	0,021863	0,22981	0,030762	0,959
1-Err↓	0,58175	0,059112	0,60511	0,061977	0,374
AvPrec↑	0,35235	0,018928	0,35461	0,025276	0,508
RLoss↓	0,43594	0,028378	0,43609	0,051903	0,646
HiLoss↓	1,71399	0,049734	1,75785	0,05971	0,114
DERISI					
HLoss↓	0,11343	0,004653	0,1156	0,003934	0,203
SAcc↑	0,05081	0,015253	0,05121	0,016177	0,889
Acc↑	0,23746	0,011478	0,22238	0,019343	0,074
1-Err↓	0,53571	0,030371	0,59518	0,071934	0,051
AvPrec↑	0,34534	0,011179	0,33477	0,032371	0,241
RLoss↓	0,44276	0,012367	0,44581	0,021293	0,760
HiLoss↓	1,78757	0,087188	1,81209	0,133362	0,575
EISEN					
HLoss↓	0,11636	0,004	0,12185	0,005235	0,022
SAcc↑	0,06646	0,023619	0,06013	0,023342	0,678
Acc↑	0,24469	0,0267	0,23039	0,02268	0,114
1-Err↓	0,59276	0,055831	0,65043	0,037617	0,034
AvPrec↑	0,35502	0,027071	0,35008	0,015022	0,721
RLoss↓	0,43272	0,028273	0,44302	0,02288	0,508
HiLoss↓	1,8308	0,124863	1,88077	0,164984	0,169
GASCH1					
HLoss↓	0,11004	0,003345	0,11265	0,003123	0,022
SAcc↑	0,07892	0,015097	0,07114	0,024428	0,173
Acc↑	0,25386	0,020121	0,23784	0,029505	0,059
1-Err↓	0,54665	0,043376	0,60108	0,059212	0,059
AvPrec↑	0,36747	0,015976	0,35559	0,011462	0,093
RLoss↓	0,43371	0,016405	0,44621	0,017387	0,093
HiLoss↓	1,78802	0,052924	1,77775	0,131985	0,646
GASCH2					
HLoss↓	0,11128	0,003392	0,11373	0,003176	0,028
SAcc↑	0,06923	0,01599	0,06354	0,01386	0,515
Acc↑	0,23339	0,015425	0,22097	0,018331	0,169
1-Err↓	0,52893	0,046224	0,56967	0,057566	0,114
AvPrec↑	0,34668	0,019542	0,34164	0,027311	0,386
RLoss↓	0,45391	0,012497	0,45787	0,005714	0,169
HiLoss↓	1,75958	0,079107	1,82059	0,091317	0,114
SPO					
HLoss↓	0,11189	0,002729	0,11368	0,004865	0,074
SAcc↑	0,06154	0,015507	0,0599	0,010164	0,593
Acc↑	0,23195	0,019524	0,22865	0,016845	0,445
1-Err↓	0,56885	0,048809	0,58333	0,048831	0,594
AvPrec↑	0,34453	0,029979	0,3376	0,012588	0,241
RLoss↓	0,44463	0,026461	0,44167	0,032209	0,799
HiLoss↓	1,77835	0,067785	1,77964	0,130786	0,959

Tabela A.14: Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado visão simples Nível 2.

	20	30	40	50		
CELLCYCLE						
Medida	Média	Média	Média	Média	FRI	WIL
HLoss↓	0,12563	0,11519	0,12089	0,12556	0,001	ABEF
SAcc↑	0,06747	0,07335	0,05843	0,06955	0,463	
Acc↑	0,24043	0,2375	0,23239	0,23787	0,615	
1-Err↓	0,66541	0,64763	0,64538	0,68341	0,021	F
AvPrec↑	0,34961	0,3412	0,36097	0,34434	0,293	
RLoss↓	0,44175	0,43856	0,43661	0,45049	0,516	
HiLoss↓	2,01337	1,79883	1,94968	1,97582	0,002	ADE
CHURCH						
HLoss↓	0,11661	0,11519	0,11916	0,11789	0,018	DE
SAcc↑	0,07499	0,07335	0,07828	0,0754	0,780	
Acc↑	0,23631	0,2375	0,23612	0,2416	0,989	
1-Err↓	0,63942	0,64763	0,65994	0,63249	0,232	
AvPrec↑	0,33967	0,3412	0,33962	0,34474	0,923	
RLoss↓	0,43799	0,43856	0,44121	0,44076	0,782	
HiLoss↓	1,80859	1,79883	1,81968	1,83606	0,668	
DERISI						
HLoss↓	0,12072	0,12184	0,12037	0,12072	0,989	
SAcc↑	0,06483	0,07139	0,06318	0,06279	0,608	
Acc↑	0,2358	0,2341	0,23631	0,23614	0,948	
1-Err↓	0,66242	0,65711	0,64149	0,61937	0,198	
AvPrec↑	0,33282	0,34285	0,33607	0,34121	0,840	
RLoss↓	0,43161	0,43199	0,43709	0,43723	0,753	
HiLoss↓	1,85684	1,81826	1,78537	1,90023	0,033	AF
EISEN						
HLoss↓	0,12879	0,1291	0,1261	0,12885	0,840	
SAcc↑	0,06967	0,07155	0,06903	0,06207	0,401	
Acc↑	0,25531	0,25221	0,25126	0,25407	0,840	
1-Err↓	0,66125	0,65484	0,65924	0,65362	0,591	
AvPrec↑	0,36932	0,36826	0,3689	0,3692	0,948	
RLoss↓	0,42785	0,4294	0,43144	0,43517	0,840	
HiLoss↓	2,03795	2,04688	1,98287	2,04617	0,356	
GASCH1						
HLoss↓	0,12111	0,12145	0,12165	0,12399	0,888	
SAcc↑	0,08062	0,074	0,08634	0,07355	0,672	
Acc↑	0,27543	0,2716	0,27965	0,26838	0,724	
1-Err↓	0,64938	0,64078	0,63824	0,65255	0,591	
AvPrec↑	0,38236	0,38116	0,38597	0,37775	0,668	
RLoss↓	0,41006	0,42396	0,42612	0,42465	0,392	
HiLoss↓	1,96474	1,95623	1,97702	1,99542	0,431	
GASCH2						
HLoss↓	0,12226	0,12288	0,12327	0,12294	0,752	
SAcc↑	0,06965	0,07336	0,07337	0,06929	0,806	
Acc↑	0,23383	0,24208	0,24715	0,23525	0,516	
1-Err↓	0,63897	0,65683	0,64103	0,63529	0,339	
AvPrec↑	0,35068	0,34536	0,34714	0,34627	0,971	
RLoss↓	0,43447	0,43332	0,43986	0,43434	0,948	
HiLoss↓	1,95915	1,94902	1,96334	1,96485	0,929	
SPO						
HLoss↓	0,12166	0,12027	0,11927	0,11956	0,840	
SAcc↑	0,06861	0,07315	0,07978	0,07319	0,570	
Acc↑	0,2368	0,23345	0,24536	0,23998	0,373	
1-Err↓	0,63793	0,65186	0,65725	0,67292	0,118	
AvPrec↑	0,34861	0,34434	0,34853	0,33783	0,494	
RLoss↓	0,41965	0,42723	0,42948	0,4213	0,668	
HiLoss↓	1,94663	1,90732	1,91026	1,85817	0,339	

Tabela A.15: Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado multivisão Nível 2.

	20	30	40	50		
CELLCYCLE						
Medida	Média	Média	Média	Média	FRI	WIL
HLoss↓	0,11143	0,11326	0,11368	0,11311	0,043	B
SAcc↑	0,05922	0,06375	0,06087	0,06661	0,194	
Acc↑	0,22364	0,23057	0,22565	0,23097	0,266	
1-Err↓	0,55705	0,56321	0,55457	0,5513	0,965	
AvPrec↑	0,35208	0,34867	0,34247	0,34968	0,516	
RLoss↓	0,46014	0,45376	0,45357	0,44815	0,430	
HiLoss↓	1,76271	1,83797	1,82648	1,87317	0,084	
CHURCH						
HLoss↓	0,11219	0,11275	0,11203	0,11344	0,796	
SAcc↑	0,07244	0,07778	0,07122	0,07777	0,861	
Acc↑	0,22073	0,22981	0,2253	0,22975	0,668	
1-Err↓	0,60547	0,59158	0,59158	0,58175	0,494	
AvPrec↑	0,34529	0,35461	0,35235	0,34049	0,062	
RLoss↓	0,43594	0,43634	0,43609	0,43757	0,451	
HiLoss↓	1,71399	1,73483	1,7557	1,72016	0,696	
DERISI						
HLoss↓	0,11391	0,11412	0,11343	0,11385	0,916	
SAcc↑	0,04998	0,05081	0,05121	0,04958	0,579	
Acc↑	0,22543	0,22269	0,23746	0,23529	0,266	
1-Err↓	0,57511	0,58086	0,53571	0,58946	0,022	BD
AvPrec↑	0,33982	0,33666	0,34534	0,33972	0,989	
RLoss↓	0,44581	0,44717	0,44276	0,44451	0,840	
HiLoss↓	1,81209	1,8339	1,8313	1,78757	0,323	
EISEN						
HLoss↓	0,11749	0,11733	0,11636	0,11679	0,976	
SAcc↑	0,06646	0,06455	0,06453	0,06138	0,953	
Acc↑	0,23668	0,24469	0,23427	0,23461	0,782	
1-Err↓	0,6333	0,60423	0,60295	0,59276	0,772	
AvPrec↑	0,34821	0,35326	0,35273	0,35502	0,840	
RLoss↓	0,43272	0,44164	0,43983	0,43622	0,782	
HiLoss↓	1,8308	1,84345	1,86374	1,86064	0,711	
GASCH1						
HLoss↓	0,11004	0,11067	0,11015	0,11066	0,985	
SAcc↑	0,07892	0,07155	0,07114	0,07196	0,459	
Acc↑	0,25386	0,2434	0,24137	0,2456	0,145	
1-Err↓	0,56218	0,55776	0,54665	0,56098	0,899	
AvPrec↑	0,36371	0,3667	0,36747	0,36547	0,971	
RLoss↓	0,43623	0,43547	0,43371	0,43441	0,896	
HiLoss↓	1,77775	1,82935	1,78672	1,83546	0,164	
GASCH2						
HLoss↓	0,11136	0,11128	0,11193	0,11154	0,468	
SAcc↑	0,06638	0,0623	0,06923	0,06354	0,756	
Acc↑	0,22921	0,23339	0,23077	0,23229	0,971	
1-Err↓	0,55664	0,55416	0,52893	0,55542	0,834	
AvPrec↑	0,3461	0,34476	0,34431	0,34668	0,840	
RLoss↓	0,45492	0,45643	0,45404	0,45391	0,896	
HiLoss↓	1,75958	1,79624	1,83077	1,84791	0,100	
SPO						
HLoss↓	0,11189	0,11196	0,1119	0,11309	0,160	
SAcc↑	0,05906	0,0603	0,05906	0,06154	0,537	
Acc↑	0,22466	0,22435	0,22204	0,23195	0,724	
1-Err↓	0,5779	0,56885	0,57296	0,57085	0,840	
AvPrec↑	0,34453	0,34416	0,33699	0,33868	0,540	
RLoss↓	0,45194	0,44463	0,44167	0,44337	0,896	
HiLoss↓	1,77835	1,79448	1,82962	1,86522	0,210	

Tabela A.16: Média e desvio padrão das abordagens semissupervisionada multivisão *versus* semissupervisionada visão simples *versus* abordagem supervisionada Nível 2.

	Co-Training		Self-Training		Supervisionado			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,1135	0,004988	0,12732	0,005322	0,11943	0,002891	0,000	ABC
SAcc↑	0,04255	0,008025	0,06955	0,018537	0,06413	0,019342	0,002	AB
Acc↑	0,22865	0,014005	0,24043	0,017629	0,2377	0,021431	0,061	
1-Err↓	0,5513	0,03659	0,63923	0,066227	0,52879	0,032814	0,000	AC
AvPrec↑	0,35208	0,02044	0,35333	0,019747	0,38953	0,017381	0,008	BC
RLoss↓	0,44815	0,016829	0,44667	0,019982	0,44646	0,019025	1,000	
HiLoss↓	1,86755	0,103502	2,04582	0,053768	1,93689	0,061442	0,001	AC
CHURCH								
HLoss↓	0,119	0,007231	0,13175	0,003565	0,11509	0,005026	0,002	AC
SAcc↑	0,05287	0,028316	0,07828	0,015646	0,02941	0,012489	0,001	ABC
Acc↑	0,22981	0,030762	0,2416	0,015599	0,19952	0,014033	0,001	BC
1-Err↓	0,58175	0,059112	0,73126	0,073419	0,61525	0,047886	0,006	AC
AvPrec↑	0,35461	0,025276	0,34474	0,016844	0,37755	0,014932	0,006	BC
RLoss↓	0,43757	0,02329	0,44076	0,031989	0,39963	0,019711	0,000	BC
HiLoss↓	1,85373	0,085071	2,05531	0,067732	1,77092	0,091093	0,000	ABC
DERISI								
HLoss↓	0,11567	0,004972	0,12687	0,004929	0,11921	0,004568	0,001	AC
SAcc↑	0,04102	0,022853	0,07139	0,013124	0,05284	0,012591	0,014	AC
Acc↑	0,23293	0,017749	0,23631	0,014108	0,24509	0,006153	0,061	
1-Err↓	0,53571	0,030371	0,61937	0,040607	0,54996	0,023636	0,001	AC
AvPrec↑	0,34534	0,011179	0,34285	0,024147	0,40288	0,005262	0,000	BC
RLoss↓	0,44451	0,01656	0,43409	0,015561	0,36608	0,006928	0,000	ABC
HiLoss↓	1,90611	0,068172	2,03733	0,064311	1,85887	0,076045	0,000	AC
EISEN								
HLoss↓	0,1191	0,004702	0,12885	0,008703	0,12553	0,00525	0,007	AB
SAcc↑	0,04431	0,019792	0,07155	0,014255	0,0601	0,021745	0,002	AB
Acc↑	0,23568	0,019964	0,25531	0,020637	0,24065	0,027755	0,407	
1-Err↓	0,59276	0,055831	0,65362	0,076064	0,54209	0,053618	0,045	C
AvPrec↑	0,35502	0,027071	0,36932	0,027607	0,38005	0,027019	0,082	
RLoss↓	0,43272	0,028273	0,42785	0,016676	0,44289	0,026012	0,497	
HiLoss↓	1,90362	0,132683	2,04688	0,073425	1,9835	0,059084	0,050	AC
GASCHI								
HLoss↓	0,11004	0,003345	0,1242	0,003479	0,11952	0,004222	0,000	ABC
SAcc↑	0,04094	0,014461	0,08634	0,011993	0,05064	0,019084	0,001	AC
Acc↑	0,23417	0,019438	0,27965	0,014794	0,27168	0,020832	0,001	AB
1-Err↓	0,54665	0,043376	0,63955	0,054531	0,5425	0,031789	0,001	AC
AvPrec↑	0,36747	0,015976	0,38597	0,012237	0,40687	0,017909	0,001	ABC
RLoss↓	0,43547	0,013893	0,42701	0,010808	0,41875	0,016838	0,082	
HiLoss↓	1,78802	0,052924	2,00164	0,035515	1,95413	0,06125	0,000	AB
GASCH2								
HLoss↓	0,11321	0,004238	0,12713	0,006023	0,11862	0,004348	0,000	ABC
SAcc↑	0,05053	0,015414	0,07337	0,018401	0,06839	0,013388	0,061	
Acc↑	0,23229	0,021718	0,24715	0,02001	0,23366	0,016746	0,150	
1-Err↓	0,52893	0,046224	0,63529	0,060432	0,55412	0,024786	0,006	AC
AvPrec↑	0,34668	0,019542	0,35068	0,016669	0,38262	0,012604	0,002	BC
RLoss↓	0,45423	0,024559	0,44151	0,01999	0,45665	0,016654	0,273	
HiLoss↓	1,85134	0,126037	2,04359	0,057516	1,90211	0,06405	0,001	AC
SPO								
HLoss↓	0,11539	0,004652	0,12641	0,004739	0,11951	0,004361	0,000	AC
SAcc↑	0,04341	0,022987	0,07978	0,01325	0,06941	0,025435	0,016	A
Acc↑	0,22463	0,014847	0,24536	0,018582	0,23755	0,015414	0,082	
1-Err↓	0,56885	0,048809	0,63793	0,063479	0,53528	0,027494	0,001	AC
AvPrec↑	0,34453	0,029979	0,34861	0,023284	0,38706	0,016968	0,002	BC
RLoss↓	0,45493	0,021047	0,44569	0,020579	0,39496	0,018052	0,000	BC
HiLoss↓	1,86522	0,075442	2,01992	0,080197	1,91851	0,066359	0,002	AC

Tabela A.17: Média e o desvio padrão dos métodos BR, CC e LP supervisionados Nível 3.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,05181	0,002807	0,05673	0,001901	0,05606	0,001821	0,000	AB
SAcc↑	0,02985	0,013628	0,03886	0,011744	0,03313	0,00729	0,442	
Acc↑	0,18808	0,014797	0,19484	0,019818	0,18454	0,013016	0,273	
1-Err↓	0,53989	0,043114	0,53989	0,043114	0,59347	0,042075	0,000	BC
AvPrec↑	0,26913	0,02046	0,26913	0,02046	0,2548	0,020252	0,000	BC
RLoss↓	0,49715	0,033089	0,49358	0,01495	0,51634	0,024251	0,150	
HiLoss↓	2,10217	0,130259	2,28878	0,038752	2,30878	0,047602	0,000	AB
CHURCH								
HLoss↓	0,0496	0,003188	0,05585	0,001337	0,05364	0,001196	0,002	ABC
SAcc↑	0,01311	0,006634	0,03155	0,010923	0,03277	0,016785	0,025	AB
Acc↑	0,12372	0,012763	0,18174	0,017756	0,17306	0,016219	0,001	AB
1-Err↓	0,6456	0,025113	0,56657	0,051979	0,7464	0,031044	0,000	ABC
AvPrec↑	0,20339	0,011833	0,24776	0,020818	0,16404	0,010597	0,000	ABC
RLoss↓	0,45572	0,022189	0,46085	0,020558	0,45442	0,031309	0,122	
HiLoss↓	1,88982	0,148478	2,29662	0,038561	2,28965	0,030296	0,001	AB
DERISI								
HLoss↓	0,05201	0,001935	0,05721	0,001777	0,0551	0,001053	0,000	ABC
SAcc↑	0,0201	0,009557	0,02831	0,012437	0,02831	0,007852	0,479	
Acc↑	0,16394	0,011771	0,17984	0,014896	0,17009	0,008697	0,273	
1-Err↓	0,55745	0,020784	0,55745	0,020784	0,60749	0,020312	0,000	BC
AvPrec↑	0,25676	0,010939	0,25676	0,010939	0,24282	0,013779	0,000	BC
RLoss↓	0,42331	0,018828	0,42488	0,018454	0,50359	0,022765	0,000	BC
HiLoss↓	1,97744	0,064977	2,31747	0,06583	2,31702	0,050537	0,001	AB
EISEN								
HLoss↓	0,0582	0,002368	0,06253	0,003175	0,0624	0,002675	0,007	AB
SAcc↑	0,02342	0,008462	0,03989	0,01634	0,0323	0,011729	0,029	A
Acc↑	0,20281	0,015183	0,19956	0,020502	0,18701	0,016753	0,497	
1-Err↓	0,55288	0,063127	0,55288	0,063127	0,59721	0,05532	0,000	BC
AvPrec↑	0,26216	0,027966	0,26216	0,027966	0,24952	0,026017	0,000	BC
RLoss↓	0,4803	0,021759	0,48482	0,021762	0,51382	0,032741	0,020	BC
HiLoss↓	2,20206	0,117543	2,35207	0,074389	2,34824	0,084552	0,020	AB
GASCHI								
HLoss↓	0,05208	0,00229	0,05458	0,001906	0,05564	0,001978	0,007	AB
SAcc↑	0,02291	0,011275	0,04624	0,015889	0,03807	0,010222	0,001	AB
Acc↑	0,20569	0,017412	0,22798	0,016194	0,19938	0,013497	0,061	
1-Err↓	0,58427	0,038767	0,55278	0,029268	0,62645	0,035208	0,000	BC
AvPrec↑	0,26828	0,014743	0,29411	0,016468	0,23432	0,019374	0,000	ABC
RLoss↓	0,47024	0,018584	0,46696	0,017998	0,50244	0,018631	0,007	BC
HiLoss↓	2,1358	0,073524	2,26187	0,087367	2,30359	0,036343	0,001	AB
GASCH2								
HLoss↓	0,05229	0,001117	0,05576	0,002436	0,05474	0,002219	0,007	AB
SAcc↑	0,03628	0,009722	0,03668	0,014772	0,03628	0,009722	0,895	
Acc↑	0,18129	0,014923	0,19187	0,020182	0,18223	0,007549	0,500	
1-Err↓	0,54439	0,030692	0,54439	0,030692	0,59819	0,025429	0,000	BC
AvPrec↑	0,27544	0,016859	0,27544	0,016859	0,2604	0,015197	0,000	BC
RLoss↓	0,49917	0,017286	0,49722	0,005691	0,49019	0,03235	0,273	
HiLoss↓	2,07904	0,057694	2,25137	0,05413	2,27262	0,06569	0,000	
SPO								
HLoss↓	0,05056	0,00259	0,05584	0,001871	0,0535	0,001922	0,001	ABC
SAcc↑	0,02979	0,013508	0,03969	0,01808	0,03804	0,020489	0,146	
Acc↑	0,17453	0,020719	0,19356	0,015615	0,18935	0,018951	0,150	
1-Err↓	0,53287	0,041668	0,53287	0,041668	0,58618	0,034857	0,000	BC
AvPrec↑	0,27206	0,02275	0,27206	0,02275	0,25741	0,023542	0,000	BC
RLoss↓	0,45729	0,011485	0,46102	0,010958	0,50656	0,014408	0,000	ABC
HiLoss↓	2,05991	0,11161	2,26746	0,070167	2,28236	0,087881	0,001	AB

Tabela A.18: Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado visão simples Nível 3.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,05286	0,002271	0,0553	0,002598	0,05633	0,001757	0,007	AB
SAcc↑	0,01554	0,007449	0,03765	0,009011	0,03193	0,013947	0,001	AB
Acc↑	0,18183	0,014781	0,18638	0,012517	0,18311	0,01067	0,497	
1-Err↓	0,68829	0,038104	0,68343	0,060611	0,82986	0,0152	0,001	BC
AvPrec↑	0,22375	0,015895	0,21821	0,014962	0,20591	0,010007	0,007	BC
RLoss↓	0,49902	0,048792	0,48871	0,04945	0,46636	0,027624	0,497	
HiLoss↓	2,13736	0,109584	2,20162	0,094055	2,30675	0,030526	0,014	BC
CHURCH								
HLoss↓	0,04908	0,001969	0,05578	0,001684	0,05426	0,001824	0,002	AB
SAcc↑	0,01065	0,006735	0,03319	0,013024	0,03319	0,01425	0,002	AB
Acc↑	0,14889	0,034365	0,18804	0,021099	0,17152	0,012783	0,027	AC
1-Err↓	0,66037	0,038534	0,81155	0,041107	0,82385	0,039447	0,001	AB
AvPrec↑	0,19492	0,030442	0,21451	0,023103	0,19593	0,017621	0,497	
RLoss↓	0,48977	0,031673	0,48942	0,025099	0,48422	0,054528	0,407	
HiLoss↓	1,86356	0,124882	2,28719	0,053619	2,29457	0,039847	0,000	AB
DERISI								
HLoss↓	0,05132	0,001732	0,05494	0,002754	0,05506	0,001676	0,005	AB
SAcc↑	0,01025	0,004832	0,03077	0,010796	0,03036	0,006759	0,000	AB
Acc↑	0,1572	0,02011	0,17875	0,010651	0,17	0,012834	0,082	
1-Err↓	0,67352	0,060037	0,678	0,089339	0,82857	0,023954	0,002	BC
AvPrec↑	0,20032	0,018522	0,20452	0,011807	0,19531	0,011761	0,497	
RLoss↓	0,47891	0,027768	0,4755	0,025005	0,4849	0,027244	0,905	
HiLoss↓	1,96067	0,194636	2,18411	0,095402	2,31664	0,040595	0,000	ABC
EISEN								
HLoss↓	0,06038	0,002602	0,06043	0,002942	0,06344	0,002369	0,021	BC
SAcc↑	0,01583	0,008032	0,03228	0,017248	0,02976	0,009933	0,032	B
Acc↑	0,1997	0,018642	0,191	0,022685	0,18525	0,012036	0,202	
1-Err↓	0,72447	0,040842	0,6928	0,045985	0,83343	0,025566	0,000	BC
AvPrec↑	0,24061	0,021685	0,2301	0,020673	0,21197	0,014065	0,006	B
RLoss↓	0,48406	0,03833	0,48689	0,030418	0,48455	0,024407	0,794	
HiLoss↓	2,24632	0,108446	2,25576	0,093383	2,38814	0,105404	0,001	BC
GASCH1								
HLoss↓	0,05235	0,002363	0,05359	0,001202	0,05591	0,001435	0,002	BC
SAcc↑	0,018	0,005844	0,03356	0,009984	0,03806	0,01581	0,001	AB
Acc↑	0,20119	0,016843	0,21912	0,010481	0,19515	0,015195	0,150	AC
1-Err↓	0,69185	0,033878	0,68861	0,05045	0,82696	0,024853	0,001	BC
AvPrec↑	0,23743	0,016578	0,24868	0,014201	0,21521	0,013894	0,006	BC
RLoss↓	0,46775	0,054219	0,46542	0,046158	0,45361	0,019943	0,905	
HiLoss↓	2,13252	0,087485	2,22876	0,03655	2,30237	0,051886	0,000	ABC
GASCH2								
HLoss↓	0,05234	0,002247	0,05426	0,002685	0,05533	0,001791	0,046	B
SAcc↑	0,01834	0,006479	0,03588	0,018141	0,03669	0,01374	0,002	AB
Acc↑	0,17077	0,019312	0,1897	0,017412	0,18256	0,010631	0,301	
1-Err↓	0,66745	0,034874	0,65768	0,065428	0,8211	0,023469	0,001	BC
AvPrec↑	0,21396	0,020279	0,21897	0,01776	0,20664	0,012565	0,273	
RLoss↓	0,50346	0,051899	0,50942	0,019064	0,45918	0,033349	0,025	BC
HiLoss↓	2,05987	0,097043	2,15196	0,072403	2,28728	0,068521	0,000	BC
SPO								
HLoss↓	0,05058	0,003282	0,05462	0,003094	0,05416	0,001914	0,020	AB
SAcc↑	0,01447	0,009787	0,04382	0,015236	0,03388	0,015561	0,000	AB
Acc↑	0,16665	0,008085	0,19148	0,011928	0,1869	0,013259	0,002	AB
1-Err↓	0,67836	0,029264	0,69455	0,087484	0,80985	0,021136	0,002	BC
AvPrec↑	0,20506	0,018019	0,21887	0,009435	0,21188	0,011658	0,045	
RLoss↓	0,4945	0,012921	0,49376	0,013174	0,47137	0,03019	0,025	C
HiLoss↓	2,00499	0,106986	2,18725	0,127204	2,30595	0,07495	0,000	ABC

Tabela A.19: Média e o desvio padrão dos métodos BR, CC e LP semissupervisionado multivisão Nível 3.

	BR		CC		LP			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,04752	0,001	0,04789	0,001591	0,05612	0,001616	0,001	BC
SAcc↑	0,01186	0,006517	0,01718	0,010358	0,03191	0,013645	0,033	B
Acc↑	0,15974	0,013712	0,16864	0,013004	0,17872	0,020959	0,202	
1-Err↓	0,5906	0,058982	0,56198	0,042503	0,82618	0,028966	0,001	BC
AvPrec↑	0,20609	0,028945	0,2129	0,010674	0,20215	0,011858	0,273	
RLoss↓	0,50296	0,061497	0,49502	0,021123	0,49217	0,044322	1,000	
HiLoss↓	1,88253	0,092487	1,92963	0,07573	2,30304	0,060716	0,001	BC
CHURCH								
HLoss↓	0,04705	0,00164	0,04861	0,003776	0,05406	0,001751	0,002	BC
SAcc↑	0,13177	0,015318	0,03196	0,021061	0,03566	0,010607	0,000	AB
Acc↑	0,1346	0,012645	0,17968	0,01418	0,17627	0,012004	0,001	AB
1-Err↓	0,61657	0,030845	0,58832	0,054494	0,82141	0,035873	0,000	BC
AvPrec↑	0,18385	0,022418	0,21854	0,022779	0,20185	0,015182	0,027	A
RLoss↓	0,47728	0,015898	0,47842	0,029002	0,48503	0,035222	0,497	
HiLoss↓	1,77922	0,104555	1,86927	0,174256	2,29661	0,049637	0,001	BC
DERISI								
HLoss↓	0,04713	0,001653	0,04829	0,00167	0,05537	0,00118	0,000	ABC
SAcc↑	0,00861	0,005618	0,02217	0,012877	0,02668	0,011505	0,003	AB
Acc↑	0,14783	0,016106	0,17016	0,024322	0,17088	0,013789	0,061	
1-Err↓	0,59845	0,042489	0,57508	0,057035	0,83556	0,026609	0,000	BC
AvPrec↑	0,19361	0,016993	0,21487	0,016438	0,19477	0,014476	0,067	
RLoss↓	0,48907	0,023894	0,49153	0,027321	0,50382	0,037016	0,061	
HiLoss↓	1,85758	0,111009	1,94139	0,146744	2,32196	0,06658	0,001	BC
EISEN								
HLoss↓	0,05312	0,001601	0,0538	0,002012	0,06452	0,001762	0,000	BC
SAcc↑	0,0133	0,009661	0,02216	0,017192	0,0323	0,016175	0,020	B
Acc↑	0,17033	0,017829	0,17726	0,022442	0,1819	0,018583	0,741	
1-Err↓	0,64409	0,035442	0,62125	0,071243	0,51617	0,01786	0,000	BC
AvPrec↑	0,20965	0,015564	0,22017	0,018598	0,20627	0,018563	0,301	
RLoss↓	0,49987	0,017831	0,49594	0,029633	0,49756	0,034319	0,905	
HiLoss↓	1,96767	0,061323	1,95941	0,059558	2,36484	0,158301	0,001	BC
GASCH1								
HLoss↓	0,04666	0,001128	0,04721	0,001837	0,05637	0,001769	0,000	BC
SAcc↑	0,01719	0,007158	0,02415	0,015725	0,03969	0,01416	0,003	B
Acc↑	0,17975	0,016011	0,18735	0,029922	0,195	0,021651	0,150	
1-Err↓	0,60516	0,024402	0,56056	0,065953	0,82736	0,019323	0,001	BC
AvPrec↑	0,22712	0,015155	0,23347	0,026518	0,21588	0,019262	0,202	
RLoss↓	0,50301	0,022965	0,48856	0,023671	0,4905	0,062292	0,670	
HiLoss↓	1,82204	0,074864	1,91661	0,100887	2,29918	0,044515	0,000	BC
GASCH2								
HLoss↓	0,04657	0,001018	0,04704	0,000952	0,05535	0,00113	0,000	BC
SAcc↑	0,01629	0,009186	0,02527	0,00917	0,03341	0,009141	0,001	ABC
Acc↑	0,15922	0,019977	0,18287	0,023486	0,18154	0,013832	0,025	AB
1-Err↓	0,57663	0,049824	0,54157	0,06116	0,82518	0,022856	0,000	BC
AvPrec↑	0,20171	0,0182	0,21658	0,019958	0,20249	0,013274	0,273	
RLoss↓	0,52169	0,012168	0,52355	0,013755	0,48594	0,062957	0,045	
HiLoss↓	1,82805	0,037645	1,89687	0,076781	2,28771	0,037601	0,000	ABC
SPO								
HLoss↓	0,04707	0,001567	0,04886	0,002588	0,05402	0,001494	0,000	ABC
SAcc↑	0,01199	0,0069	0,02399	0,008686	0,03475	0,00997	0,000	ABC
Acc↑	0,15279	0,012307	0,17772	0,020282	0,18496	0,014628	0,001	AB
1-Err↓	0,59779	0,064859	0,55974	0,033493	0,81688	0,026913	0,001	BC
AvPrec↑	0,20041	0,016155	0,22516	0,018052	0,20916	0,013889	0,002	AC
RLoss↓	0,50541	0,020755	0,50906	0,020006	0,47589	0,030482	0,061	
HiLoss↓	1,88457	0,084956	1,99423	0,089009	2,30342	0,054265	0,000	BC

Tabela A.20: Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado visão simples Nível 3.

	50		75		
CELLCYCLE					
Medida	Média	Des. Padrão	Média	Des. Padrão	WIL
HLoss↓	0,05286	0,002271	0,0541	0,002073	0,114
SAcc↑	0,03765	0,009011	0,03193	0,013947	0,236
Acc↑	0,18638	0,012517	0,18384	0,016282	0,445
1-Err↓	0,68829	0,038104	0,68343	0,060611	0,878
AvPrec↑	0,22375	0,015895	0,21821	0,014962	0,333
RLoss↓	0,4937	0,027821	0,46636	0,027624	0,059
HiLoss↓	2,13736	0,109584	2,14598	0,091022	0,878
CHURCH					
HLoss↓	0,04908	0,001969	0,04987	0,002137	0,575
SAcc↑	0,03318	0,013024	0,03319	0,01425	0,674
Acc↑	0,18804	0,021099	0,17574	0,019005	0,086
1-Err↓	0,66037	0,038534	0,66694	0,047518	0,721
AvPrec↑	0,21451	0,023103	0,20261	0,023137	0,139
RLoss↓	0,48691	0,0436	0,48422	0,054528	0,959
HiLoss↓	1,86356	0,124882	1,9382	0,135125	0,386
DERISI					
HLoss↓	0,05132	0,001732	0,05241	0,001767	0,139
SAcc↑	0,03077	0,010796	0,03036	0,006759	0,889
Acc↑	0,17875	0,010651	0,16981	0,010195	0,114
1-Err↓	0,67352	0,060037	0,678	0,089339	0,575
AvPrec↑	0,20452	0,011807	0,20412	0,010809	0,859
RLoss↓	0,47896	0,021035	0,4755	0,025005	0,646
HiLoss↓	1,99902	0,185581	1,96067	0,194636	0,646
EISEN					
HLoss↓	0,06038	0,002602	0,06104	0,003731	0,799
SAcc↑	0,02976	0,011186	0,03228	0,017248	0,507
Acc↑	0,1997	0,018642	0,191	0,022685	0,508
1-Err↓	0,71125	0,060914	0,6928	0,045985	0,508
AvPrec↑	0,24061	0,021685	0,2301	0,020673	0,445
RLoss↓	0,48834	0,02147	0,48406	0,03833	0,575
HiLoss↓	2,24632	0,108446	2,27728	0,141961	0,333
GASCH1					
HLoss↓	0,05235	0,002363	0,05436	0,002576	0,093
SAcc↑	0,03724	0,012864	0,03806	0,01581	0,878
Acc↑	0,21912	0,010481	0,20389	0,018311	0,139
1-Err↓	0,69185	0,033878	0,68861	0,05045	0,878
AvPrec↑	0,24868	0,014201	0,2362	0,01288	0,074
RLoss↓	0,45361	0,019943	0,4542	0,025801	0,878
HiLoss↓	2,13252	0,087485	2,19439	0,114792	0,203
GASCH2					
HLoss↓	0,05234	0,002247	0,05255	0,002398	0,575
SAcc↑	0,03669	0,01374	0,03179	0,01085	0,374
Acc↑	0,1897	0,017412	0,18291	0,016727	0,271
1-Err↓	0,66745	0,034874	0,65768	0,065428	0,959
AvPrec↑	0,21897	0,01776	0,21749	0,019079	0,799
RLoss↓	0,46695	0,020693	0,45918	0,033349	0,799
HiLoss↓	2,07749	0,110896	2,05987	0,097043	0,721
SPO					
HLoss↓	0,05058	0,003282	0,05155	0,002619	0,475
SAcc↑	0,04382	0,015236	0,031	0,015855	0,066
Acc↑	0,19148	0,011928	0,18084	0,012952	0,093
1-Err↓	0,67836	0,029264	0,69077	0,056534	0,721
AvPrec↑	0,21887	0,009435	0,21035	0,01152	0,059
RLoss↓	0,47421	0,021174	0,47137	0,03019	0,721
HiLoss↓	2,00499	0,106986	2,0769	0,132487	0,074

Tabela A.21: Média e o desvio padrão dos percentuais iniciais de exemplos não rotulados para o método semissupervisionado multivisão Nível 3.

	50		75		
CELLCYCLE					
Medida	Média	Des. Padrão	Média	Des. Padrão	WIL
HLoss↓	0,04752	0,001	0,04859	0,000913	0,028
SAcc↑	0,03191	0,013645	0,02946	0,010205	0,635
Acc↑	0,17872	0,020959	0,17569	0,01379	0,445
1-Err↓	0,56198	0,042503	0,6237	0,064657	0,047
AvPrec↑	0,2129	0,010674	0,21005	0,019602	0,575
RLoss↓	0,49502	0,021123	0,49217	0,044322	0,799
HiLoss↓	1,88253	0,092487	1,90951	0,154628	0,445
CHURCH					
HLoss↓	0,04705	0,00164	0,04837	0,002892	0,203
SAcc↑	0,03566	0,010607	0,13177	0,015318	0,005
Acc↑	0,17637	0,022699	0,17968	0,01418	0,508
1-Err↓	0,58832	0,054494	0,63665	0,101786	0,114
AvPrec↑	0,21854	0,022779	0,21664	0,016488	0,646
RLoss↓	0,48005	0,020309	0,47728	0,015898	0,575
HiLoss↓	1,77922	0,104555	1,83079	0,098027	0,386
DERISI					
HLoss↓	0,04713	0,001653	0,04876	0,002019	0,028
SAcc↑	0,02627	0,00973	0,02668	0,011505	1,000
Acc↑	0,17088	0,013789	0,16863	0,013058	0,575
1-Err↓	0,57508	0,057035	0,58367	0,065036	0,721
AvPrec↑	0,21487	0,016438	0,20356	0,017714	0,285
RLoss↓	0,48907	0,023894	0,49247	0,038183	0,646
HiLoss↓	1,85758	0,111009	1,89084	0,140527	0,575
EISEN					
HLoss↓	0,05312	0,001601	0,05514	0,002744	0,009
SAcc↑	0,03167	0,009885	0,0323	0,016175	0,959
Acc↑	0,1819	0,018583	0,17396	0,019859	0,386
1-Err↓	0,51617	0,01786	0,64348	0,06397	0,005
AvPrec↑	0,21521	0,025412	0,22017	0,018598	0,721
RLoss↓	0,49594	0,029633	0,49756	0,034319	0,799
HiLoss↓	1,95941	0,059558	2,00569	0,168673	0,767
GASCH1					
HLoss↓	0,04666	0,001128	0,04841	0,000987	0,005
SAcc↑	0,03969	0,01416	0,03642	0,012684	0,646
Acc↑	0,195	0,021651	0,18638	0,011199	0,386
1-Err↓	0,56056	0,065953	0,58959	0,057622	0,285
AvPrec↑	0,23347	0,026518	0,2278	0,022238	0,878
RLoss↓	0,48856	0,023671	0,4905	0,062292	0,799
HiLoss↓	1,82204	0,074864	1,90014	0,134645	0,139
GASCH2					
HLoss↓	0,04657	0,001018	0,04768	0,001518	0,009
SAcc↑	0,03341	0,009141	0,03138	0,015581	0,635
Acc↑	0,18287	0,023486	0,18154	0,013832	0,646
1-Err↓	0,54157	0,06116	0,57045	0,063363	0,333
AvPrec↑	0,21658	0,019958	0,21491	0,022791	0,721
RLoss↓	0,49967	0,03914	0,48594	0,062957	0,646
HiLoss↓	1,82805	0,037645	1,83699	0,058763	0,767
SPO					
HLoss↓	0,04707	0,001567	0,04828	0,001991	0,047
SAcc↑	0,03475	0,00997	0,02978	0,00777	0,203
Acc↑	0,18496	0,014628	0,17684	0,014055	0,037
1-Err↓	0,55974	0,033493	0,58003	0,08485	0,508
AvPrec↑	0,22516	0,018052	0,20645	0,017504	0,007
RLoss↓	0,48946	0,033752	0,47589	0,030482	0,139
HiLoss↓	1,88457	0,084956	1,95486	0,140751	0,139

Tabela A.22: Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado visão simples Nível 3.

	20	30	40	50		
CELLCYCLE						
Medida	Média	Média	Média	Média	FRI	WIL
HLoss↓	0,05286	0,05301	0,05367	0,0535	0,724	
SAcc↑	0,03765	0,03765	0,03029	0,03644	0,129	
Acc↑	0,18371	0,18275	0,18521	0,18638	0,868	
1-Err↓	0,68343	0,69196	0,69402	0,69204	0,592	
AvPrec↑	0,22375	0,21385	0,21587	0,21821	0,421	
RLoss↓	0,49296	0,4937	0,46636	0,46899	0,160	
HiLoss↓	2,14324	2,13736	2,14598	2,16759	0,971	
CHURCH						
HLoss↓	0,05016	0,04987	0,04931	0,04908	0,753	
SAcc↑	0,03319	0,0291	0,03278	0,03319	0,382	
Acc↑	0,18016	0,17555	0,1764	0,18804	0,145	
1-Err↓	0,66037	0,68455	0,67019	0,66694	0,451	
AvPrec↑	0,20426	0,20023	0,20121	0,21451	0,017	E
RLoss↓	0,4898	0,48422	0,48942	0,48925	0,564	
HiLoss↓	1,91482	1,91069	1,87918	1,86356	0,451	
DERISI						
HLoss↓	0,05259	0,05257	0,05132	0,05241	0,145	
SAcc↑	0,03077	0,02749	0,03077	0,02708	0,921	
Acc↑	0,17751	0,16984	0,17875	0,1762	0,923	
1-Err↓	0,67759	0,70217	0,68498	0,67352	0,293	
AvPrec↑	0,20419	0,20412	0,20452	0,20331	0,644	
RLoss↓	0,48105	0,4755	0,48024	0,47896	0,840	
HiLoss↓	1,96067	2,04267	1,99902	2,0025	0,615	
EISEN						
HLoss↓	0,06156	0,06043	0,06109	0,06038	0,339	
SAcc↑	0,02976	0,02851	0,02848	0,03228	0,920	
Acc↑	0,1997	0,191	0,18983	0,18886	0,840	
1-Err↓	0,6928	0,69665	0,71625	0,70295	0,596	
AvPrec↑	0,24061	0,23011	0,2312	0,22846	0,564	
RLoss↓	0,48836	0,48689	0,48455	0,48406	0,845	
HiLoss↓	2,27728	2,27676	2,25576	2,24632	0,631	
GASCH1						
HLoss↓	0,05402	0,05235	0,05359	0,05407	0,537	
SAcc↑	0,03357	0,03602	0,03806	0,03724	0,898	
Acc↑	0,20831	0,20733	0,21912	0,20467	0,197	
1-Err↓	0,69684	0,68861	0,69185	0,69679	0,743	
AvPrec↑	0,23978	0,2362	0,24868	0,23563	0,615	
RLoss↓	0,46118	0,4542	0,46646	0,45361	0,451	
HiLoss↓	2,18822	2,13252	2,2148	2,19439	0,218	
GASCH2						
HLoss↓	0,05314	0,05234	0,05236	0,05255	0,868	
SAcc↑	0,03588	0,03669	0,03139	0,03182	0,566	
Acc↑	0,1897	0,18709	0,18079	0,18133	0,451	
1-Err↓	0,70335	0,66745	0,65768	0,66664	0,066	
AvPrec↑	0,21897	0,21817	0,21018	0,21749	0,696	
RLoss↓	0,46145	0,48988	0,47061	0,45918	0,668	
HiLoss↓	2,11162	2,08481	2,08108	2,05987	0,323	
SPO						
HLoss↓	0,05416	0,05083	0,05122	0,05058	0,008	ABC
SAcc↑	0,03514	0,03184	0,0335	0,04382	0,146	
Acc↑	0,1869	0,17711	0,18619	0,19148	0,112	
1-Err↓	0,7069	0,6937	0,67836	0,69077	0,816	
AvPrec↑	0,21188	0,21035	0,21581	0,21887	0,086	
RLoss↓	0,48167	0,47421	0,4872	0,47137	0,266	
HiLoss↓	2,09132	2,05875	2,09258	2,00499	0,339	

Tabela A.23: Média dos percentuais de incorporação por iteração de novos exemplos rotulados para o método semissupervisionado multivisão Nível 3.

	20	30	40	50		
CELLCYCLE						
Medida	Média	Média	Média	Média	FRI	WIL
HLoss↓	0,04789	0,04779	0,04752	0,04796	0,856	
SAcc↑	0,03069	0,03191	0,02782	0,02906	0,670	
Acc↑	0,17815	0,17872	0,17749	0,17821	0,696	
1-Err↓	0,57621	0,59346	0,58689	0,56198	0,587	
AvPrec↑	0,20886	0,21005	0,20904	0,2129	0,896	
RLoss↓	0,49502	0,49217	0,49925	0,5093	0,782	
HiLoss↓	1,90951	1,94002	1,88253	1,96224	0,145	
CHURCH						
HLoss↓	0,04758	0,04729	0,04705	0,04804	0,531	
SAcc↑	0,03525	0,13177	0,02951	0,03566	0,000	ADE
Acc↑	0,17527	0,17637	0,16655	0,17968	0,392	
1-Err↓	0,58832	0,61657	0,61821	0,6178	0,396	
AvPrec↑	0,21854	0,21288	0,20217	0,21664	0,356	
RLoss↓	0,47842	0,48392	0,48171	0,47728	0,615	
HiLoss↓	1,79929	1,82838	1,78242	1,77922	0,473	
DERISI						
HLoss↓	0,04746	0,04735	0,04713	0,04818	0,026	F
SAcc↑	0,02667	0,02668	0,02462	0,02544	0,566	
Acc↑	0,17088	0,16825	0,16799	0,16863	0,782	
1-Err↓	0,57508	0,58367	0,59679	0,62189	0,334	
AvPrec↑	0,21487	0,20356	0,20733	0,21247	0,782	
RLoss↓	0,49807	0,49247	0,48907	0,49875	0,218	
HiLoss↓	1,85758	1,89084	1,89333	1,87378	0,971	
EISEN						
HLoss↓	0,0538	0,05422	0,05368	0,05312	0,985	
SAcc↑	0,03167	0,03107	0,0323	0,0298	0,730	
Acc↑	0,1819	0,17629	0,17726	0,17635	0,840	
1-Err↓	0,63137	0,51617	0,64348	0,62125	0,000	ADE
AvPrec↑	0,21419	0,2131	0,21163	0,22017	0,948	
RLoss↓	0,49739	0,49756	0,49594	0,50218	0,896	
HiLoss↓	1,9824	1,96767	1,95941	1,99548	0,916	
GASCH1						
HLoss↓	0,04674	0,04666	0,04721	0,04784	0,073	
SAcc↑	0,03561	0,02989	0,03276	0,03969	0,169	
Acc↑	0,18735	0,18043	0,18549	0,195	0,131	
1-Err↓	0,56959	0,57815	0,56056	0,58027	0,682	
AvPrec↑	0,23347	0,22103	0,23101	0,23076	0,293	
RLoss↓	0,49472	0,49085	0,48856	0,4905	0,896	
HiLoss↓	1,82204	1,85961	1,899	1,9636	0,056	
GASCH2						
HLoss↓	0,04727	0,04657	0,04673	0,04676	0,497	
SAcc↑	0,03058	0,03341	0,033	0,03138	0,618	
Acc↑	0,18287	0,17931	0,18154	0,17515	0,668	
1-Err↓	0,58306	0,57045	0,54157	0,55421	0,566	
AvPrec↑	0,21658	0,21491	0,21372	0,21655	0,965	
RLoss↓	0,50761	0,49967	0,48594	0,51113	0,564	
HiLoss↓	1,86101	1,82805	1,83699	1,83047	0,929	
SPO						
HLoss↓	0,0471	0,04707	0,0476	0,04789	0,254	
SAcc↑	0,02978	0,02935	0,03475	0,02977	0,496	
Acc↑	0,18097	0,18259	0,18496	0,18151	0,896	
1-Err↓	0,59327	0,56761	0,59733	0,55974	0,392	
AvPrec↑	0,20647	0,22516	0,20916	0,20985	0,036	D
RLoss↓	0,47589	0,50046	0,50204	0,49032	0,339	
HiLoss↓	1,93796	1,88457	1,95576	1,97602	0,096	

Tabela A.24: Média e desvio padrão das abordagens semissupervisionada multivisão *versus* semissupervisionada visão simples *versus* abordagem supervisionada Nível 3.

	Co-Training		Self-Training		Supervisionado			
CELLCYCLE								
Medida	Média	Des. Padrão	Média	Des. Padrão	Média	Des. Padrão	FRI	WIL
HLoss↓	0,04789	0,001591	0,0553	0,002598	0,05673	0,001901	0,000	BC
SAcc↑	0,01718	0,010358	0,03765	0,009011	0,03886	0,011744	0,001	AB
Acc↑	0,16864	0,013004	0,18638	0,012517	0,19484	0,019818	0,007	AB
1-Err↓	0,56198	0,042503	0,68343	0,060611	0,53989	0,043114	0,001	AC
AvPrec↑	0,2129	0,010674	0,21821	0,014962	0,26913	0,02046	0,000	BC
RLoss↓	0,49502	0,021123	0,48871	0,04945	0,49358	0,01495	0,670	
HiLoss↓	1,92963	0,07573	2,20162	0,094055	2,28878	0,038752	0,000	ABC
CHURCH								
HLoss↓	0,04861	0,003776	0,05578	0,001684	0,05585	0,001337	0,002	AB
SAcc↑	0,03196	0,021061	0,03319	0,013024	0,03155	0,010923	0,779	
Acc↑	0,17968	0,01418	0,18804	0,021099	0,18174	0,017756	0,407	
1-Err↓	0,58832	0,054494	0,81155	0,041107	0,56657	0,051979	0,001	AC
AvPrec↑	0,21854	0,022779	0,21451	0,023103	0,24776	0,020818	0,020	C
RLoss↓	0,47842	0,029002	0,48942	0,025099	0,46085	0,020558	0,045	C
HiLoss↓	1,86927	0,174256	2,28719	0,053619	2,29662	0,038561	0,001	AB
DERISI								
HLoss↓	0,04829	0,00167	0,05494	0,002754	0,05721	0,001777	0,000	ABC
SAcc↑	0,02217	0,012877	0,03077	0,010796	0,02831	0,012437	0,313	
Acc↑	0,17016	0,024322	0,17875	0,010651	0,17984	0,014896	0,905	
1-Err↓	0,57508	0,057035	0,678	0,089339	0,55745	0,020784	0,002	AC
AvPrec↑	0,21487	0,016438	0,20452	0,011807	0,25676	0,010939	0,001	BC
RLoss↓	0,49153	0,027321	0,4755	0,025005	0,42488	0,018454	0,000	BC
HiLoss↓	1,94139	0,146744	2,18411	0,095402	2,31747	0,06583	0,000	ABC
EISEN								
HLoss↓	0,0538	0,002012	0,06043	0,002942	0,06253	0,003175	0,000	AB
SAcc↑	0,02216	0,017192	0,03228	0,017248	0,03989	0,01634	0,110	
Acc↑	0,17726	0,022442	0,191	0,022685	0,19956	0,020502	0,150	
1-Err↓	0,62125	0,071243	0,6928	0,045985	0,55288	0,063127	0,001	ABC
AvPrec↑	0,22017	0,018598	0,2301	0,020673	0,26216	0,027966	0,020	BC
RLoss↓	0,49594	0,029633	0,48689	0,030418	0,48482	0,021762	0,670	
HiLoss↓	1,95941	0,059558	2,25576	0,093383	2,35207	0,074389	0,000	ABC
GASCH1								
HLoss↓	0,04721	0,001837	0,05359	0,001202	0,05458	0,001906	0,000	AB
SAcc↑	0,02415	0,015725	0,03356	0,009984	0,04624	0,015889	0,020	B
Acc↑	0,18735	0,029922	0,21912	0,010481	0,22798	0,016194	0,007	AB
1-Err↓	0,56056	0,065953	0,68861	0,05045	0,55278	0,029268	0,006	AC
AvPrec↑	0,23347	0,026518	0,24868	0,014201	0,29411	0,016468	0,000	BC
RLoss↓	0,48856	0,023671	0,46542	0,046158	0,46696	0,017998	0,045	B
HiLoss↓	1,91661	0,100887	2,22876	0,03655	2,26187	0,087367	0,000	AB
GASCH2								
HLoss↓	0,04704	0,000952	0,05426	0,002685	0,05576	0,002436	0,001	AB
SAcc↑	0,02527	0,00917	0,03588	0,018141	0,03668	0,014772	0,179	
Acc↑	0,18287	0,023486	0,1897	0,017412	0,19187	0,020182	0,209	
1-Err↓	0,54157	0,06116	0,65768	0,065428	0,54439	0,030692	0,002	AC
AvPrec↑	0,21658	0,019958	0,21897	0,01776	0,27544	0,016859	0,001	BC
RLoss↓	0,52355	0,013755	0,50942	0,019064	0,49722	0,005691	0,007	B
HiLoss↓	1,89687	0,076781	2,15196	0,072403	2,25137	0,05413	0,000	ABC
SPO								
HLoss↓	0,04886	0,002588	0,05462	0,003094	0,05584	0,001871	0,001	AB
SAcc↑	0,02399	0,008686	0,04382	0,015236	0,03969	0,01808	0,010	AB
Acc↑	0,17772	0,020282	0,19148	0,011928	0,19356	0,015615	0,407	
1-Err↓	0,55974	0,033493	0,69455	0,087484	0,53287	0,041668	0,000	AC
AvPrec↑	0,22516	0,018052	0,21887	0,009435	0,27206	0,02275	0,001	BC
RLoss↓	0,50906	0,020006	0,49376	0,013174	0,46102	0,010958	0,001	BC
HiLoss↓	1,99423	0,089009	2,18725	0,127204	2,26746	0,070167	0,001	AB

Tabela A.25: Média dos Melhores Resultados do Aprendizado Supervisionado nos Três Níveis da Classificação Hierárquica Multirrotulo.

	BR	CC	LP	
CELLCYCLE				
Medida	Média	Média	Média	Nível
HLoss↓	0,30026	0,30531	0,3037	1
HLoss↓	0,11943	0,12989	0,13385	2
HLoss↓	0,05181	0,05673	0,05606	3
SAcc↑	0,24901	0,3223	0,33002	1
SAcc↑	0,06413	0,09205	0,06741	2
SAcc↑	0,02985	0,03886	0,03313	3
Acc↑	0,40954	0,44507	0,44479	1
Acc↑	0,2377	0,264	0,23521	2
Acc↑	0,18808	0,19484	0,18454	3
1-Err↓	0,4384	0,44578	0,44448	1
1-Err↓	0,52879	0,54485	0,58481	2
1-Err↓	0,53989	0,53989	0,58481	3
AvPrec↑	0,71052	0,7059	0,71522	1
AvPrec↑	0,38953	0,3896	0,37742	2
AvPrec↑	0,26913	0,26913	0,37742	3
RLoss↓	0,33575	0,34119	0,3084	1
RLoss↓	0,44646	0,44342	0,46386	2
RLoss↓	0,49715	0,49358	0,51634	3
HiLoss↓				1
HiLoss↓	1,93689	2,06468	2,13534	2
HiLoss↓	2,10217	2,28878	2,30878	3
CHURCH				
HLoss↓	0,30638	0,31114	0,31488	1
HLoss↓	0,11509	0,13171	0,13608	2
HLoss↓	0,0496	0,05585	0,05364	3
SAcc↑	0,24453	0,32408	0,31168	1
SAcc↑	0,02941	0,06803	0,07039	2
SAcc↑	0,01311	0,03155	0,03277	3
Acc↑	0,38425	0,43505	0,42231	1
Acc↑	0,19952	0,23951	0,22568	2
Acc↑	0,12372	0,18174	0,17306	3
1-Err↓	0,46699	0,44938	0,45594	1
1-Err↓	0,61525	0,49243	0,73737	2
1-Err↓	0,6456	0,56657	0,7464	3
AvPrec↑	0,70038	0,70836	0,71029	1
AvPrec↑	0,37755	0,3801	0,27702	2
AvPrec↑	0,20339	0,24776	0,16404	3
RLoss↓	0,34635	0,3406	0,30542	1
RLoss↓	0,39963	0,4041	0,41615	2
RLoss↓	0,45572	0,46085	0,45442	3
HiLoss↓				1
HiLoss↓	1,77092	2,07373	2,16219	2
HiLoss↓	1,88982	2,29662	2,28965	3
DERISI				
HLoss↓	0,30267	0,31367	0,31199	1
HLoss↓	0,11921	0,13543	0,13608	2
HLoss↓	0,05201	0,05721	0,0551	3
SAcc↑	0,27231	0,32489	0,31989	1
SAcc↑	0,05284	0,06689	0,05284	2
SAcc↑	0,0201	0,02831	0,02831	3
Acc↑	0,38703	0,4342	0,43058	1
Acc↑	0,24509	0,24489	0,21856	2
Acc↑	0,16394	0,17984	0,17009	3
1-Err↓	0,45978	0,46145	0,45404	1
1-Err↓	0,54996	0,55006	0,60249	2
1-Err↓	0,55745	0,55745	0,60749	3
AvPrec↑	0,70525	0,70424	0,71307	1
AvPrec↑	0,40288	0,40198	0,36261	2
AvPrec↑	0,25676	0,25676	0,24282	3
RLoss↓	0,33754	0,34012	0,30332	1
RLoss↓	0,36608	0,36737	0,44479	2
RLoss↓	0,42331	0,42488	0,50359	3
HiLoss↓				1
HiLoss↓	1,85887	2,11484	2,18656	2
HiLoss↓	1,97744	2,31747	2,31702	3
EISEN				

HLoss↓	0,30013	0,31047	0,31771	1
HLoss↓	0,12553	0,13329	0,13631	2
HLoss↓	0,0582	0,06253	0,0624	3
SAcc↑	0,25836	0,31918	0,30587	1
SAcc↑	0,0601	0,08106	0,06138	2
SAcc↑	0,02342	0,03989	0,0323	3
Acc↑	0,4272	0,44433	0,42693	1
Acc↑	0,24065	0,25923	0,2239	2
Acc↑	0,20281	0,19956	0,18701	3
1-Err↓	0,42617	0,44645	0,45971	1
1-Err↓	0,54209	0,54212	0,59338	2
1-Err↓	0,55288	0,55288	0,59721	3
AvPrec↑	0,72413	0,7141	0,7108	1
AvPrec↑	0,38005	0,38011	0,36936	2
AvPrec↑	0,26216	0,26216	0,24952	3
RLoss↓	0,32024	0,33163	0,31464	1
RLoss↓	0,44289	0,44808	0,45627	2
RLoss↓	0,4803	0,48482	0,51382	3
HiLoss↓				1
HiLoss↓	1,9835	2,09308	2,18685	2
HiLoss↓	2,20206	2,35207	2,34824	3
GASCH1				
HLoss↓	0,2871	0,28876	0,30027	1
HLoss↓	0,11952	0,12461	0,13004	2
HLoss↓	0,05208	0,05458	0,05564	3
SAcc↑	0,28715	0,36127	0,33833	1
SAcc↑	0,05064	0,08879	0,07972	2
SAcc↑	0,02291	0,04624	0,03807	3
Acc↑	0,43584	0,48529	0,45527	1
Acc↑	0,27168	0,28311	0,25531	2
Acc↑	0,20569	0,22798	0,19938	3
1-Err↓	0,41648	0,40915	0,43568	1
1-Err↓	0,5425	0,50823	0,62802	2
1-Err↓	0,58427	0,55278	0,62645	3
AvPrec↑	0,7285	0,73358	0,72122	1
AvPrec↑	0,40687	0,412	0,35792	2
AvPrec↑	0,26828	0,29411	0,23432	3
RLoss↓	0,31438	0,3125	0,30317	1
RLoss↓	0,41875	0,43361	0,45408	2
RLoss↓	0,47024	0,46696	0,50244	3
HiLoss↓				1
HiLoss↓	1,95413	2,03597	2,1018	2
HiLoss↓	2,1358	2,26187	2,30359	3
GASCH2				
HLoss↓	0,29865	0,31104	0,30376	1
HLoss↓	0,11862	0,13091	0,13229	2
HLoss↓	0,05229	0,05576	0,05474	3
SAcc↑	0,25991	0,3276	0,33083	1
SAcc↑	0,06839	0,08636	0,07167	2
SAcc↑	0,03628	0,03668	0,03628	3
Acc↑	0,39899	0,43786	0,44447	1
Acc↑	0,23366	0,24908	0,23552	2
Acc↑	0,18129	0,19187	0,18223	3
1-Err↓	0,44535	0,44705	0,43722	1
1-Err↓	0,55412	0,5497	0,60712	2
1-Err↓	0,54439	0,54439	0,59819	3
AvPrec↑	0,70782	0,70594	0,71957	1
AvPrec↑	0,38262	0,38264	0,37072	2
AvPrec↑	0,27544	0,27544	0,2604	3
RLoss↓	0,34154	0,34445	0,29989	1
RLoss↓	0,45665	0,45857	0,45833	2
RLoss↓	0,49917	0,49722	0,49019	3
HiLoss↓				1
HiLoss↓	1,90211	2,04241	2,10998	2
HiLoss↓	2,07904	2,25137	2,27262	3
SPO				
HLoss↓	0,31093	0,31905	0,30161	1
HLoss↓	0,11951	0,13288	0,13215	2
HLoss↓	0,05056	0,05584	0,0535	3
SAcc↑	0,22607	0,3051	0,33523	1
SAcc↑	0,06941	0,08391	0,06941	2
SAcc↑	0,02979	0,03969	0,03804	3

Acc↑	0,39556	0,4196	0,44945	1
Acc↑	0,23755	0,24698	0,23727	2
Acc↑	0,17453	0,19356	0,18935	3
1-Err↓	0,45759	0,46175	0,43608	1
1-Err↓	0,53528	0,53534	0,58611	2
1-Err↓	0,53287	0,53287	0,58618	3
AvPrec↑	0,69804	0,69435	0,72258	1
AvPrec↑	0,38706	0,3871	0,37511	2
AvPrec↑	0,27206	0,27206	0,25741	3
RLoss↓	0,35414	0,35816	0,29985	1
RLoss↓	0,39496	0,39438	0,45543	2
RLoss↓	0,45729	0,46102	0,50656	3
HiLoss↓				1
HiLoss↓	1,91851	2,05205	2,12233	2
HiLoss↓	2,05991	2,26746	2,28236	3

Tabela A.26: Média dos Melhores Resultados do Aprendizado Semissupervisionado Visão Simples nos Três Níveis da Classificação Hierárquica Multirrotulo.

	BR	CC	LP	
CELLCYCLE				
Medida	Média	Média	Média	Nível
HLoss↓	0,30496	0,31492	0,30986	1
HLoss↓	0,12089	0,12732	0,13504	2
HLoss↓	0,05286	0,0553	0,05633	3
SAcc↑	0,24821	0,30836	0,31896	1
SAcc↑	0,03469	0,06955	0,06373	2
SAcc↑	0,01554	0,03765	0,03193	3
Acc↑	0,39311	0,43025	0,43749	1
Acc↑	0,23239	0,24043	0,22939	2
Acc↑	0,18183	0,18638	0,18311	3
1-Err↓	0,46374	0,46215	0,45596	1
1-Err↓	0,64538	0,63923	0,78724	2
1-Err↓	0,68829	0,68343	0,82986	3
AvPrec↑	0,69531	0,69614	0,70814	1
AvPrec↑	0,36097	0,35333	0,32935	2
AvPrec↑	0,22375	0,21821	0,20591	3
RLoss↓	0,35597	0,35287	0,31771	1
RLoss↓	0,44834	0,44667	0,43661	2
RLoss↓	0,49902	0,48871	0,46636	3
HiLoss↓				1
HiLoss↓	1,94968	2,04582	2,146	2
HiLoss↓	2,13736	2,20162	2,30675	3
CHURCH				
HLoss↓	0,30126	0,30898	0,31528	1
HLoss↓	0,11519	0,13175	0,13606	2
HLoss↓	0,04908	0,05578	0,05426	3
SAcc↑	0,2617	0,32736	0,31005	1
SAcc↑	0,02695	0,07828	0,07203	2
SAcc↑	0,01065	0,03319	0,03319	3
Acc↑	0,37801	0,44163	0,42203	1
Acc↑	0,21945	0,2416	0,22791	2
Acc↑	0,14889	0,18804	0,17152	3
1-Err↓	0,46576	0,4461	0,45717	1
1-Err↓	0,63249	0,73126	0,78075	2
1-Err↓	0,66037	0,81155	0,82385	3
AvPrec↑	0,70224	0,71115	0,70984	1
AvPrec↑	0,33088	0,34474	0,32774	2
AvPrec↑	0,19492	0,21451	0,19593	3
RLoss↓	0,34416	0,33536	0,30551	1
RLoss↓	0,43799	0,44076	0,43856	2
RLoss↓	0,48977	0,48942	0,48422	3
HiLoss↓				1
HiLoss↓	1,79883	2,05531	2,16833	2
HiLoss↓	1,86356	2,28719	2,29457	3
DERISI				
HLoss↓	0,30984	0,31204	0,31414	1
HLoss↓	0,12037	0,12687	0,13615	2

HLoss↓	0,05132	0,05494	0,05506	3
SAcc↑	0,24647	0,31627	0,3121	1
SAcc↑	0,0278	0,07139	0,05162	2
SAcc↑	0,01025	0,03077	0,03036	3
Acc↑	0,38053	0,43208	0,42578	1
Acc↑	0,23536	0,23631	0,22521	2
Acc↑	0,1572	0,17875	0,17	3
1-Err↓	0,47661	0,4594	0,45609	1
1-Err↓	0,64149	0,61937	0,80716	2
1-Err↓	0,67352	0,678	0,82857	3
AvPrec↑	0,69351	0,70433	0,7111	1
AvPrec↑	0,33363	0,34285	0,31705	2
AvPrec↑	0,20032	0,20452	0,19531	3
RLoss↓	0,34866	0,33552	0,30558	1
RLoss↓	0,43161	0,43409	0,43199	2
RLoss↓	0,47891	0,4755	0,4849	3
HiLoss↓				1
HiLoss↓	1,78537	2,03733	2,17303	2
HiLoss↓	1,96067	2,18411	2,31664	3
EISEN				
HLoss↓	0,31091	0,31743	0,31327	1
HLoss↓	0,1261	0,12885	0,13745	2
HLoss↓	0,06038	0,06043	0,06344	3
SAcc↑	0,24443	0,29513	0,31725	1
SAcc↑	0,03731	0,07155	0,06579	2
SAcc↑	0,01583	0,03228	0,02976	3
Acc↑	0,39995	0,43393	0,43737	1
Acc↑	0,2415	0,25531	0,24253	2
Acc↑	0,1997	0,191	0,18525	3
1-Err↓	0,45719	0,45532	0,45336	1
1-Err↓	0,65924	0,65362	0,77261	2
1-Err↓	0,72447	0,6928	0,83343	3
AvPrec↑	0,7044	0,70165	0,71406	1
AvPrec↑	0,3689	0,36932	0,34454	2
AvPrec↑	0,24061	0,2301	0,21197	3
RLoss↓	0,3387	0,34507	0,31384	1
RLoss↓	0,4294	0,42785	0,44037	2
RLoss↓	0,48406	0,48689	0,48455	3
HiLoss↓				1
HiLoss↓	1,98287	2,04688	2,19197	2
HiLoss↓	2,24632	2,25576	2,38814	3
GASCHI				
HLoss↓	0,29118	0,29468	0,30068	1
HLoss↓	0,12111	0,1242	0,13004	2
HLoss↓	0,05235	0,05359	0,05591	3
SAcc↑	0,27652	0,33963	0,33869	1
SAcc↑	0,04371	0,08634	0,08423	2
SAcc↑	0,018	0,03356	0,03806	3
Acc↑	0,43472	0,47052	0,4565	1
Acc↑	0,25573	0,27965	0,25693	2
Acc↑	0,20119	0,21912	0,19515	3
1-Err↓	0,43815	0,42674	0,4357	1
1-Err↓	0,63824	0,63955	0,76263	2
1-Err↓	0,69185	0,68861	0,82696	3
AvPrec↑	0,72073	0,7213	0,72028	1
AvPrec↑	0,37463	0,38597	0,35475	2
AvPrec↑	0,23743	0,24868	0,21521	3
RLoss↓	0,31427	0,32213	0,30328	1
RLoss↓	0,42842	0,42701	0,41006	2
RLoss↓	0,46775	0,46542	0,45361	3
HiLoss↓				1
HiLoss↓	1,95623	2,00164	2,09853	2
HiLoss↓	2,13252	2,22876	2,30237	3
GASCH2				
HLoss↓	0,30446	0,31123	0,30955	1
HLoss↓	0,12226	0,12713	0,13399	2
HLoss↓	0,05234	0,05426	0,05533	3
SAcc↑	0,2452	0,32437	0,32142	1
SAcc↑	0,03501	0,07337	0,06598	2
SAcc↑	0,01834	0,03588	0,03669	3
Acc↑	0,38607	0,43802	0,43565	1
Acc↑	0,22215	0,24715	0,23304	2

Acc↑	0,17077	0,1897	0,18256	3
1-Err↓	0,46577	0,45274	0,44825	1
1-Err↓	0,64787	0,63529	0,7848	2
1-Err↓	0,66745	0,65768	0,8211	3
AvPrec↑	0,69507	0,70537	0,71264	1
AvPrec↑	0,34714	0,35068	0,33054	2
AvPrec↑	0,21396	0,21897	0,20664	3
RLoss↓	0,35187	0,34175	0,30684	1
RLoss↓	0,4516	0,44151	0,43332	2
RLoss↓	0,50346	0,50942	0,45918	3
HiLoss↓				1
HiLoss↓	1,94902	2,04359	2,12425	2
HiLoss↓	2,05987	2,15196	2,28728	3
SPO				
HLoss↓	0,30823	0,31314	0,30317	1
HLoss↓	0,11927	0,12641	0,13296	2
HLoss↓	0,05058	0,05462	0,05416	3
SAcc↑	0,22653	0,3142	0,32779	1
SAcc↑	0,03259	0,07978	0,07188	2
SAcc↑	0,01447	0,04382	0,03388	3
Acc↑	0,36807	0,42771	0,44512	1
Acc↑	0,2327	0,24536	0,23948	2
Acc↑	0,16665	0,19148	0,1869	3
1-Err↓	0,4621	0,45761	0,43729	1
1-Err↓	0,65103	0,63793	0,77962	2
1-Err↓	0,67836	0,69455	0,80985	3
AvPrec↑	0,69717	0,69354	0,71936	1
AvPrec↑	0,34786	0,34861	0,33826	2
AvPrec↑	0,20506	0,21887	0,21188	3
RLoss↓	0,35286	0,3594	0,30301	1
RLoss↓	0,44554	0,44569	0,41965	2
RLoss↓	0,4945	0,49376	0,47137	3
HiLoss↓				1
HiLoss↓	1,85817	2,01992	2,11778	2
HiLoss↓	2,00499	2,18725	2,30595	3

Tabela A.27: Média dos Melhores Resultados do Aprendizado Semissupervisionado Multivisão nos Três Níveis da Classificação Hierárquica Multirrótulo.

	BR	CC	LP	
CELLCYCLE				
Medida	Média	Média	Média	Nível
HLoss↓	0,29863	0,29989	0,30403	1
HLoss↓	0,11143	0,1135	0,13539	2
HLoss↓	0,04752	0,04789	0,05612	3
SAcc↑	0,26539	0,29119	0,32917	1
SAcc↑	0,02527	0,04255	0,06661	2
SAcc↑	0,01186	0,01718	0,03191	3
Acc↑	0,38132	0,40235	0,44539	1
Acc↑	0,21525	0,22865	0,23057	2
Acc↑	0,15974	0,16864	0,17872	3
1-Err↓	0,42978	0,43757	0,44452	1
1-Err↓	0,56346	0,5513	0,78724	2
1-Err↓	0,5906	0,56198	0,82618	3
AvPrec↑	0,7156	0,70815	0,71492	1
AvPrec↑	0,34152	0,35208	0,33184	2
AvPrec↑	0,20609	0,2129	0,20215	3
RLoss↓	0,32529	0,33223	0,31103	1
RLoss↓	0,45398	0,44815	0,45357	2
RLoss↓	0,50296	0,49502	0,49217	3
HiLoss↓				1
HiLoss↓	1,76271	1,86755	2,13984	2
HiLoss↓	1,88253	1,92963	2,30304	3
CHURCH				
HLoss↓	0,30053	0,30713	0,31633	1
HLoss↓	0,11203	0,119	0,1361	2
HLoss↓	0,04705	0,04861	0,05406	3
SAcc↑	0,25232	0,29826	0,30841	1

SAcc↑	0,02327	0,05287	0,07778	2
SAcc↑	0,13177	0,03196	0,03566	3
Acc↑	0,38692	0,41337	0,41957	1
Acc↑	0,22073	0,22981	0,22975	2
Acc↑	0,1346	0,17968	0,17627	3
1-Err↓	0,45838	0,45633	0,45963	1
1-Err↓	0,6046	0,58175	0,77624	2
1-Err↓	0,61657	0,58832	0,82141	3
AvPrec↑	0,70214	0,70612	0,70763	1
AvPrec↑	0,31446	0,35461	0,33021	2
AvPrec↑	0,18385	0,21854	0,20185	3
RLoss↓	0,34732	0,33981	0,30824	1
RLoss↓	0,43594	0,43757	0,43609	2
RLoss↓	0,47728	0,47842	0,48503	3
HiLoss↓				1
HiLoss↓	1,71399	1,85373	2,16176	2
HiLoss↓	1,77922	1,86927	2,29661	3
DERISI				
HLoss↓	0,30182	0,30733	0,31279	1
HLoss↓	0,11343	0,11567	0,13661	2
HLoss↓	0,04713	0,04829	0,05537	3
SAcc↑	0,25591	0,29045	0,31292	1
SAcc↑	0,02412	0,04102	0,05121	2
SAcc↑	0,00861	0,02217	0,02668	3
Acc↑	0,36868	0,40309	0,42753	1
Acc↑	0,23746	0,23293	0,21464	2
Acc↑	0,14783	0,17016	0,17088	3
1-Err↓	0,46143	0,45694	0,45403	1
1-Err↓	0,6021	0,53571	0,80799	2
1-Err↓	0,59845	0,57508	0,83556	3
AvPrec↑	0,70418	0,70875	0,71244	1
AvPrec↑	0,33147	0,34534	0,31566	2
AvPrec↑	0,19361	0,21487	0,19477	3
RLoss↓	0,3319	0,32175	0,30315	1
RLoss↓	0,44276	0,44451	0,44581	2
RLoss↓	0,48907	0,49153	0,50382	3
HiLoss↓				1
HiLoss↓	1,78757	1,90611	2,18161	2
HiLoss↓	1,85758	1,94139	2,32196	3
EISEN				
HLoss↓	0,30203	0,3035	0,30097	1
HLoss↓	0,11636	0,1191	0,13957	2
HLoss↓	0,05312	0,0538	0,06452	3
SAcc↑	0,25389	0,28182	0,33381	1
SAcc↑	0,03035	0,04431	0,06646	2
SAcc↑	0,0133	0,02216	0,0323	3
Acc↑	0,37593	0,41245	0,45673	1
Acc↑	0,21963	0,23568	0,24469	2
Acc↑	0,17033	0,17726	0,1819	3
1-Err↓	0,40967	0,41988	0,44143	1
1-Err↓	0,63516	0,59276	0,77073	2
1-Err↓	0,64409	0,62125	0,51617	3
AvPrec↑	0,7265	0,72042	0,71253	1
AvPrec↑	0,33873	0,35502	0,34701	2
AvPrec↑	0,20965	0,22017	0,20627	3
RLoss↓	0,32219	0,32228	0,31487	1
RLoss↓	0,4429	0,43272	0,44164	2
RLoss↓	0,49987	0,49594	0,49756	3
HiLoss↓				1
HiLoss↓	1,8308	1,90362	2,20079	2
HiLoss↓	1,96767	1,95941	2,36484	3
GASCHI				
HLoss↓	0,27962	0,27842	0,30097	1
HLoss↓	0,11015	0,11004	0,1333	2
HLoss↓	0,04666	0,04721	0,05637	3
SAcc↑	0,28962	0,32897	0,33381	1
SAcc↑	0,03102	0,04094	0,07892	2
SAcc↑	0,01719	0,02415	0,03969	3
Acc↑	0,41158	0,44915	0,45673	1
Acc↑	0,23	0,23417	0,25386	2
Acc↑	0,17975	0,18735	0,195	3
1-Err↓	0,40381	0,38747	0,44143	1

1-Err↓	0,56131	0,54665	0,76833	2
1-Err↓	0,60516	0,56056	0,82736	3
AvPrec↑	0,73447	0,74247	0,71866	1
AvPrec↑	0,36547	0,36747	0,35031	2
AvPrec↑	0,22712	0,23347	0,21588	3
RLoss↓	0,30335	0,29292	0,30651	1
RLoss↓	0,43371	0,43547	0,43623	2
RLoss↓	0,50301	0,48856	0,4905	3
HiLoss↓				1
HiLoss↓	1,77775	1,78802	2,12799	2
HiLoss↓	1,82204	1,91661	2,29918	3
GASCH2				
HLoss↓	0,29509	0,29768	0,30619	1
HLoss↓	0,11128	0,11321	0,1342	2
HLoss↓	0,04657	0,04704	0,05535	3
SAcc↑	0,27498	0,30726	0,32554	1
SAcc↑	0,03174	0,05053	0,06923	2
SAcc↑	0,01629	0,02527	0,03341	3
Acc↑	0,37532	0,41541	0,43943	1
Acc↑	0,20375	0,23229	0,23339	2
Acc↑	0,15922	0,18287	0,18154	3
1-Err↓	0,43233	0,43076	0,44293	1
1-Err↓	0,55542	0,52893	0,78559	2
1-Err↓	0,57663	0,54157	0,82518	3
AvPrec↑	0,71183	0,71327	0,71436	1
AvPrec↑	0,33351	0,34668	0,3324	2
AvPrec↑	0,20171	0,21658	0,20249	3
RLoss↓	0,32932	0,32649	0,30872	1
RLoss↓	0,46456	0,45423	0,45391	2
RLoss↓	0,52169	0,52355	0,48594	3
HiLoss↓				1
HiLoss↓	1,75958	1,85134	2,12424	2
HiLoss↓	1,82805	1,89687	2,28771	3
SPO				
HLoss↓	0,30326	0,30788	0,30399	1
HLoss↓	0,11189	0,11539	0,13418	2
HLoss↓	0,04707	0,04886	0,05402	3
SAcc↑	0,25421	0,2782	0,33233	1
SAcc↑	0,02845	0,04341	0,06154	2
SAcc↑	0,01199	0,02399	0,03475	3
Acc↑	0,37444	0,38932	0,44621	1
Acc↑	0,22149	0,22463	0,23195	2
Acc↑	0,15279	0,17772	0,18496	3
1-Err↓	0,45222	0,46091	0,44022	1
1-Err↓	0,57085	0,56885	0,78828	2
1-Err↓	0,59779	0,55974	0,81688	3
AvPrec↑	0,70673	0,69887	0,71957	1
AvPrec↑	0,33455	0,34453	0,33115	2
AvPrec↑	0,20041	0,22516	0,20916	3
RLoss↓	0,3282	0,3437	0,30463	1
RLoss↓	0,45207	0,45493	0,44167	2
RLoss↓	0,50541	0,50906	0,47589	3
HiLoss↓				1
HiLoss↓	1,77835	1,86522	2,14296	2
HiLoss↓	1,88457	1,99423	2,30342	3