



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO CIÊNCIAS DA COMPUTAÇÃO

PAULO ROGER GOMES CORDEIRO

Uma abordagem para pós-seleção de *ensembles* utilizando conceitos de Oráculo e *Multi-view Learning*

Recife

2024

PAULO ROGER GOMES CORDEIRO

Uma abordagem para pós-seleção de *ensembles* utilizando conceitos de Oráculo e *Multi-view Learning*

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Doutor em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador (a): Prof. Dr. George Darmiton da Cunha Cavalcanti

Coorientador (a): Prof. Dr. Rafael Menelau Oliveira e Cruz

Recife

2024

Catálogo na fonte
Bibliotecária: Luiza Oliveira/CRB1316

C794u Cordeiro, Paulo Roger Gomes.
Uma abordagem para a pós-seleção de *ensembles* utilizando conceitos de oráculo e *multi-view learning* / Paulo Roger Gomes Cordeiro.– 2024.
109 fls.: il.

Orientador: George Darminton da Cunha Cavalcanti.

Coorientador: Rafael Menelau Oliveira e Cruz.

Tese (Doutorado) – Universidade Federal de Pernambuco. Centro de Informática, Programa de Pós-graduação em Ciências da Computação, Recife, 2024.

Inclui referências e apêndices.

1. Pós-seleção 2. Sistemas de múltiplos classificadores. 3. Seleção dinâmica de ensembles. 4. Multi-view Learning. I. Cavalcanti, George Darminton da Cunha. II Cruz, Rafael Menelau Oliveira e III. Título.

006.31

CDD (23. ed.)

UFPE - CCEN 2024 –49

Paulo Roger Gomes Cordeiro

“Uma abordagem para pós-seleção de ensembles utilizando conceitos de Oráculo e Multi-view Learning”

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Inteligência Computacional

Aprovada em: 06/02/2024.

Orientador: Prof. Dr. George Darmiton da Cunha Cavalcanti

BANCA EXAMINADORA

Prof. Dr. Tsang Ing Ren
Centro de Informática/UFPE

Prof. Dr. Adiel Teixeira de Almeida Filho
Centro de Informática /UFPE

Prof. Dr. Rafael Ferreira Leite de Mello
Departamento de Computação / UFRPE

Prof. Dr. João Fausto Lorenzato de Oliveira
Escola Politécnica de Pernambuco / UPE

Prof. Dr. Luiz Eduardo Soares de Oliveira
Departamento de Informática / UFPR

Para João Otávio "Tavinho" Tibúrcio de Melo Cordeiro.

AGRADECIMENTOS

Primeiro, gostaria de agradecer a pessoa que mais me ajudou, acompanhou, e dividiu os bons e maus momentos durante esses anos de doutoramento. Cinthya, minha esposa, sem seu apoio nada disso seria possível. Obrigado por ser meu porto seguro nesses momentos tão complexos. E, durante todo esse processo, nasce o ser mais importante da minha vida: João “Tavinho” Otávio. No momento em que escrevo você ainda é um bebê com 1 ano e 6 meses. Contudo, é impressionante como sua presença, seu sorriso e o seu amor foram importantes durante esse período. Obrigado, meu filho.

Aos meus pais, Nalva e Célio, por todo o esforço para proporcionar uma ótima educação, tanto acadêmica quanto doméstica, assim como todo o amor. Aos meus irmãos, Guilherme, Célio Miguel, João e José, por todo o amor e carinho. Aos meus sogros, D. Lourdinha e Seu Maércio, por toda a força e apoio com João Otávio, e comigo, desde o seu nascimento. Aos meus afilhados, Rhayssa, Antônio e Mel, sou muito feliz por ter vocês em minha vida.

Aos amigos que contribuíram para esta tese, com o apoio acadêmico e emocional: Dennis Rodrigo, Janaina Camara e Maya; Aline e Dário; Marcelo “Bio” Lacerda e Caroline, Flávio Rosendo, e aos amigos da E.T.F.

Ao meu orientador, Prof. George Darmiton, por toda a paciência e apoio durante o doutorado. A forma como o senhor discute e aplica ciência é inspiradora. Ao meu co-orientador, Prof. Rafael Menelau, pelas discussões, revisões e contribuições fundamentais para o término deste trabalho.

Ao Instituto Federal de Pernambuco (IFPE), em especial ao *campus* Barreiros, pelo apoio institucional durante os anos da pesquisa. Sou muito orgulhoso por fazer parte dessa instituição, que, mesmo com todos os desafios, luta para mudar os rumos da educação em nosso país. Também agradeço ao Centro de Informática (CIn) e a UFPE. Fico orgulhoso por fazer parte dessa comunidade acadêmica.

RESUMO

Os Sistema de múltiplos classificadores (SMC) foram desenvolvidos como alternativa às abordagens tradicionais que utilizam classificadores individuais. Entre as abordagens de SMC, os algoritmos de Seleção dinâmica de *ensemble* (SDE) destacam-se por apresentarem melhor desempenho em diversos cenários. Contudo, mesmo com os resultados promissores das abordagens de SDE, dada a heterogeneidade dos cenários de aplicação, é improvável que um único algoritmo SDE seja a solução ideal em todos os casos. É diante dessa lacuna que este trabalho propõe um novo método denominado *Post-Selection Dynamic Ensemble Selection* (PS-DES). O PS-DES atua como uma estratégia de pós-seleção, avaliando e escolhendo os melhores *ensembles* selecionados por diferentes técnicas de SDE, para cada amostra apresentada. Para avaliar os *ensembles* é proposta uma nova medida denominada *Hardness-aware Oracle* (HaO). Ao contrário da avaliação tradicional, aplicada em técnicas SDE que é centrada em classificadores individuais, o HaO estende o conceito do Oráculo ao analisar como os classificadores do *ensemble* selecionado cooperam entre si. Também é proposto adicionar uma nova camada ao PS-DES, visando gerar mais diversidade ao sistema, utilizando conceitos de *Multi-view learning* (MVL) ao aplicar diversas funções de transformação nos dados originais. Essa versão do PS-DES é chamada de *Post-Selection Dynamic Ensemble Selection with Multi-View Learning* (PS-DES-MVL). Para avaliar as abordagens propostas, foram realizados experimentos em 30 bases de dados, utilizando configurações de *pools* homogêneas e heterogêneas com três classificadores-base (Perceptron, *Logistic Regression* e *Naive Bayes*), avaliando o PS-DES com quatro abordagens de SDE estabelecidas na literatura. A avaliação do PS-DES-MVL segue o mesmo arranjo experimental apresentado para o PS-DES, adicionando a aplicação de técnicas de geração de *views*. Os resultados demonstraram que o PS-DES supera ou obtém resultados similares às abordagens SDE individuais, principalmente em configurações de *pools* heterogêneas. Além disso, o PS-DES demonstrou eficiência na escolha de técnicas SDE adequadas a diferentes situações. O PS-DES-MVL, ao avaliar formas individuais de geração de *views*, obteve resultados similares ou superiores na maior parte das análises.

Palavras-chaves: pós-seleção; sistema de múltiplos classificadores; seleção dinâmica de *ensembles*; *multi-view Learning*.

ABSTRACT

Multiple classifier systems (MCS) have been developed as an alternative to traditional approaches using individual classifiers. Among SMC approaches, Dynamic Ensemble Selection (DES) algorithms stand out for better performance in various scenarios. However, even with the promising results of SDE approaches, given the heterogeneity of application scenarios, a single DES algorithm is unlikely to be the ideal solution in all cases. Given this shortcoming, this thesis proposes a new method called Post-Selection Dynamic Ensemble Selection (PS-DES). PS-DES works as a post-selection strategy, evaluating and choosing the best ensembles selected by different EDS techniques, for each sample presented. A new measure called Hardness-aware Oracle (HaO) is also proposed to evaluate the ensembles. Unlike the traditional evaluation applied to DES techniques, which focuses on individual classifiers, HaO extends the Oracle concept by analyzing how the classifiers in the selected ensemble cooperate. It is also proposed to add a new layer to PS-DES to generate more diversity in the system, using Multi-view learning (MVL) concepts by applying various transformation functions to the original data. This version of PS-DES is called Post-Selection Dynamic Ensemble Selection with Multi-View Learning (PS-DES-MVL). To evaluate the proposed approaches, experiments were carried out on 30 databases, using homogeneous and heterogeneous pool configurations with three base classifiers (Perceptron, Logistic Regression, and Naive Bayes), evaluating PS-DES with four state-of-the-art DES approaches. The evaluation of PS-DES-MVL follows the same experimental setup presented for PS-DES, adding the application of view generation techniques. The results have shown that PS-DES outperforms or achieves similar results to the individual DES approaches, especially in heterogeneous pool configurations. In addition, PS-DES demonstrated efficiency in choosing DES techniques suitable for different situations. PS-DES-MVL, when evaluating individual methods of generating views, obtained similar or superior results in most of the analyses.

Keywords: post-selection; multiple classifier system; dynamic ensemble selection; multi-view learning.

LISTA DE FIGURAS

<p>Figura 1 – As três fases de um Sistema de múltiplos classificadores. Na primeira fase, geração, é formado um <i>pool</i> $P = \{C_1, \dots, C_m\}$, com m classificadores. Na segunda fase, Seleção, os classificadores em P são selecionados para formar o <i>ensemble</i> $P' = \{C_i, C_j, \dots, C_n\}$, com n, classificadores, de forma que $P' \subset P$.</p>	24
<p>Figura 2 – Funcionamento do Bagging</p>	31
<p>Figura 3 – Diferença entre as abordagens de seleção: a) Seleção estática, b) Seleção dinâmica de classificador e c) Seleção dinâmica de ensemble.</p>	34
<p>Figura 4 – Fase 1 do PS-DES. O objetivo é gerar um <i>pool</i> de classificadores, preferencialmente diversos, e inicializar as técnicas SDE que serão utilizadas na próxima fase.</p>	55
<p>Figura 5 – Fase 2 do PS-DES, seleção e pós-seleção dos <i>ensembles</i>. O objetivo dessa fase é selecionar o <i>ensemble</i> P^{sel} com maior valor de potencial.</p>	58
<p>Figura 6 – Fase 1 do PS-DES-MVL</p>	60
<p>Figura 7 – Box-plot da frequência percentual de escolhas das técnicas SDE realizadas pelo PS-DES na Etapa 2.4, para todas as bases.</p>	82
<p>Figura 8 – Frequência percentual das escolhas das técnicas SDE realizadas pelo PS-DES na Etapa 2.4, para as bases de dados: appendicitis, cleveland, blood, transfusion, led7digit e yeast-2_vs_4.</p>	83
<p>Figura 9 – Frequência percentual das escolhas das técnicas SDE realizadas pelo PS-DES na Etapa 2.4., para as 30 bases de dados</p>	105

LISTA DE TABELAS

Tabela 1 – Relação entre pares de classificadores	36
Tabela 2 – Resumo das informações (ano de publicação, classificadores base e métodos de seleção) relacionados as técnicas de SMC utilizando classificadores-base homogêneos.	43
Tabela 3 – Lista de classificadores, e suas respectivas siglas, utilizadas nos trabalhos que utilizam <i>ensembles</i> heterogêneos.	44
Tabela 4 – Resumo das informações (ano de publicação, classificadores base e métodos de seleção) relacionados as técnicas de SMC utilizando classificadores-base heterogêneos. As técnicas que não tem nomes definidos no trabalho estão como S/N.	49
Tabela 5 – Resumo das informações (método de geração de views, classificadores-base e métodos de combinação) relacionadas aos artigos sobre Multi-View Learning e Sistemas de Múltiplos Classificadores	50
Tabela 6 – Descrição das bases de dados: quantidade de amostras, dimensões, classes, e taxa de desbalanceamento (TD).	65
Tabela 7 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro abordagens SDE (KNOP, KNORA-U, META-DES, DES-P) e o PS-Aleatório. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores NB. . . .	69
Tabela 8 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro abordagens SDE (KNOP, KNORA-U, META-DES, DES-P) e o PS-Aleatório. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores LR. . . .	69
Tabela 9 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro abordagens SDE (KNOP, KNORA-U, META-DES, DES-P) e o PS-Aleatório. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores Perceptron. . .	70

- Tabela 10 – Avaliação da acurácia média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* heterogêneo de três classificadores básicos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 71
- Tabela 11 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* heterogêneo de três classificadores básicos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 72
- Tabela 12 – Avaliação do MCC médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* heterogêneo de três classificadores básicos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 73
- Tabela 13 – Avaliação da métrica de diversidade Disagreement comparando um *pool* heterogêneo com três configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado para cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado o *pool* heterogêneo com a configuração de *pools* homogêneos apresentada na coluna. 75

Tabela 14 – Avaliação da métrica de diversidade Double-fault comparando um *pool* heterogêneo com três configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado para cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado o *pool* heterogêneo com a configuração de *pools* homogêneos apresentada na coluna. 76

Tabela 15 – Avaliação da acurácia média comparando a abordagem PS-DES utilizando um *pool* heterogêneo com três variações do PS-DES utilizando configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com *pool* heterogêneo com a configuração de *pools* homogêneos apresentada na coluna. 78

Tabela 16 – Avaliação do F-score média comparando a abordagem PS-DES utilizando um *pool* heterogêneo com três variações do PS-DES utilizando configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com *pool* heterogêneo com a configuração de *pools* homogêneo apresentada na coluna. 79

Tabela 17 – Avaliação do MCC médio comparando a abordagem PS-DES utilizando um *pool* heterogêneo com três variações do PS-DES utilizando configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com *pool* heterogêneo com a configuração de *pools* homogêneo apresentada na coluna. 80

Tabela 18 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro formas de avaliação para o potencial (Etapa 2.3): acurácia (PS-DES-ACC), F-score (PS-DES-F), MCC (PS-DES-MCC) e HaO (PS-DES-HaO). São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.	81
Tabela 19 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o t-SNE como técnica de geração de <i>view</i> . São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.	84
Tabela 20 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o PCA como técnica de geração de <i>view</i> . São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.	84
Tabela 21 – Resultado do ranking e do <i>p-valor</i> referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o <i>kernel</i> linear como técnica de geração de <i>view</i> . São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um <i>pool</i> heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.	85

- Tabela 22 – Resultado do ranking e do *p-valor* referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o *kernel* polinomial como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron. 85
- Tabela 23 – Resultado do ranking e do *p-valor* referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o *kernel* RBF como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron. 85
- Tabela 24 – Resultado do ranking e do *p-valor* referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL quando utilizado cinco funções de representação de dados: t-SNE, PCA, *Kernel* Linear (K-Linear), *Kernel* Polinomial (K-Polinomial) e *Kernel* RBF (K-RBF), de forma individual e combinadas (MVL). São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron. 86
- Tabela 25 – Avaliação da taxa de acerto média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (NB). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha *p-valor* apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 96

Tabela 26 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores-base (NB). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha <i>ranking</i> apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.	97
Tabela 27 – Avaliação do MCC comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores-base (NB). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha <i>ranking</i> apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.	98
Tabela 28 – Avaliação da taxa de acerto média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores-base (LR). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha <i>ranking</i> apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.	99
Tabela 29 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um <i>pool</i> homogêneo com 300 classificadores-base (LR). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha <i>ranking</i> apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.	100

- Tabela 30 – Avaliação do MCC comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (LR). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 101
- Tabela 31 – Avaliação da taxa de acerto média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 102
- Tabela 32 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 103
- Tabela 33 – Avaliação do MCC médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna. 104

LISTA DE ABREVIATURAS E SIGLAS

1R	OneR
AD	Árvores de decisão
ADAC	<i>Adacost</i>
BAG	Bagging
BN	<i>Bayesian network</i>
C-RF	<i>Cost-Sensitive Random Forest</i>
C-CART	<i>Cost-Sensitive CART</i>
CL	<i>Class Likelihood</i>
CSMES	<i>Cost-Sensitive Multicriteria Ensemble Selection</i>
DCP	<i>Disjunct Class Percentage</i>
DES-CV	<i>DES-competence based on label outputs and weighted majority voting</i>
DES-P	<i>Dynamic Ensemble Selection Performance</i>
DET	<i>Decision Table</i>
DF	<i>Double-Fault</i>
DG	Diversidade global
DIS	<i>Disagreement</i>
DNN	<i>Deep Neural Network</i>
DS	<i>Decision Stump</i>
ES-JSS	<i>Ensemble Selection with Joint Spectral Clustering and Structural Sparsity</i>
ET	<i>Entropy</i>
FC	<i>Fisher Classifier</i>
FH-DES	<i>Fuzzy Hyperbox Dynamic Ensemble Selection</i>
GA-EoC	<i>GA-based searching for ensemble combination</i>
GB	<i>Gaussian Bayes</i>
GP	<i>General-Purpose</i>

GPC	<i>Gaussian Process Classifier</i>
HaO	<i>Hardness-aware Oracle</i>
HBoost	<i>Heterogeneous Booting-based ensemble classifier</i>
HDEC	<i>Heterogeneous Dynamic Ensemble Classifier</i>
HESCA	<i>The Heterogeneous Ensembles of Standard Classification Algorithms</i>
HF2HM	<i>Hierarchical fusion of homogeneous and heterogeneous multi-classifiers</i>
HT	<i>Hoeffding Tree</i>
IH	<i>Instance Hardness</i>
JRIP	<i>RIPPER rule learning</i>
k*	<i>KStar</i>
kDN	<i>k-Disagreeing Neighbors</i>
k-NN	<i>k-Nearest Neighbors</i>
KNOP	<i>K-Nearest Output Profile</i>
KNORA-E	<i>K-Nearest-Oracles Eliminate</i>
KNORA-U	<i>K-Nearest-Oracles Union</i>
KNP-DES	<i>KNORA-distance dynamic ensemble selection</i>
KP	<i>KNORA-PERCENTAGE</i>
KWV	<i>Kohavi–Wolpert Variance</i>
LC	<i>Logistic Classifier</i>
LDA	<i>Linear Discriminant Analysis</i>
LDC	<i>Linear Discriminant Classifier</i>
LR	<i>Logistic Regression</i>
MCC	<i>Matthews Correlation Coefficient</i>
MDEP	<i>Margin & Diversity based ordering Ensemble Pruning</i>
MLDE	<i>Multi-Label classification with Dynamic Ensemble learning</i>
MLP	<i>Multilayer Perceptron</i>
MLS	<i>Multi-Layer Selector</i>

MVL	<i>Multi-view learning</i>
NB	<i>Naive Bayes</i>
NMC	<i>Nearest Mean Classifier</i>
PART	<i>PART Rule Learning Algorithm</i>
PCA	<i>Principal Component Analysis</i>
PS-DES-MVL	<i>Post-Selection Dynamic Ensemble Selection with Multi-View Learning</i>
PS-DES	<i>Post-Selection Dynamic Ensemble Selection</i>
QDA	<i>Quadratic Discriminant Analysis</i>
QDC	<i>Quadratic Discriminant Classifiers</i>
QP	Questão de pesquisa
RBF	<i>Radial Basis Function Network</i>
RdC	Região de competência
RF	<i>Random Forest</i>
RSM	<i>Random Subspace Method</i>
RT	RepTree
RTF	<i>Rotation Forest</i>
SDC	Seleção dinâmica de classificador
SDE	Seleção dinâmica de <i>ensemble</i>
SDIN	Seleção dinâmica
SEST	Seleção estática
SF	<i>Selective Fusion</i>
SGB	<i>Stochastic Gradient Boosting</i>
SMC	Sistema de múltiplos classificadores
SMO	<i>Sequential Minimal Optimization</i>
SVM	<i>Support Vector Machine</i>
SVMQ	<i>Support Vector Machine with Quadratic Kernel</i>
TBAG	<i>Trimmed Bagging</i>

TD	Taxa de desbalanceamento
t-SNE	<i>t-Distributed Stochastic Neighbor Embedding</i>
WMCRP	<i>Weighted Multi-classifiers Framework Based on Random Projection</i>

LISTA DE SÍMBOLOS

θ	Região de competência
δ	Competência de um classificador
α	Nível de confiança
T	Base de treinamento
$DSEL$	Base de validação
C	Classificador
P	<i>Pool</i>
P'	<i>Ensemble</i>
k	Tamanho da Região de Competência
m	Quantidade de classificadores em um <i>pool</i>
n	Quantidade de classificadores em um <i>ensemble</i>
i	Índice de um classificador em um <i>ensemble</i>
j	Índice de um classificador em um <i>ensemble</i>
v	Quantidade de <i>views</i>
b	Quantidade de <i>baggs</i>

SUMÁRIO

1	INTRODUÇÃO	23
1.1	MOTIVAÇÃO	25
1.2	CONTRIBUIÇÕES	27
1.3	ESTRUTURA DA TESE	28
2	FUNDAMENTAÇÃO TEÓRICA	29
2.1	GERAÇÃO DO <i>POOL</i> DE CLASSIFICADORES	29
2.1.1	<i>Multi-view learning</i>	30
2.2	SELEÇÃO DOS CLASSIFICADORES	33
2.3	COMBINAÇÃO DOS CLASSIFICADORES	35
2.4	DIVERSIDADE	36
2.4.1	Métricas pareadas	36
2.4.2	Métricas não-pareadas	37
2.5	ORÁCULO	38
2.6	<i>INSTANCE HARDNESS</i>	38
3	TRABALHOS RELACIONADOS	40
3.1	TRABALHOS COM CLASSIFICADORES HOMOGÊNEOS	40
3.2	TRABALHOS DE SISTEMA DE MÚLTIPLOS CLASSIFICADORES COM CLASSIFICADORES HETEROGÊNEOS	43
3.3	TRABALHOS SOBRE MULTI-VIEW LEARNING	50
3.4	DISCUSSÃO SOBRE OS TRABALHOS APRESENTADOS	50
4	MÉTODO PROPOSTO	52
4.1	<i>HARDNESS-AWARE ORACLE</i>	52
4.2	<i>Post-Selection Dynamic Ensemble Selection</i>	53
4.2.1	PS-DES - Fase 1: Geração do <i>pool</i> e preparação das técnicas SDE	54
4.2.2	PS-DES - Fase 2: Seleção e Pós-seleção dinâmica	55
4.2.3	PS-DES - Fase 3: Combinação	57
4.3	<i>Post-Selection Dynamic Ensemble Selection with Multi-View Learning</i>	58
4.3.1	Fase 1: Geração das <i>views</i> e <i>pools</i>	59
4.3.2	Fase 2: Seleção de <i>view</i>	60
4.3.3	Fases 3: seleção e pós-seleção, e Fase 4: combinação	62

5	ARRANJO EXPERIMENTAL E RESULTADOS	63
5.1	ANÁLISES PROPOSTAS	63
5.2	BASES DE DADOS	64
5.3	PARÂMETROS DOS ALGORITMOS UTILIZADOS	66
5.4	MÉTRICAS PARA AVALIAÇÃO DE DESEMPENHO	67
5.5	ANÁLISE ESTATÍSTICA	68
5.6	ANÁLISE 1: AVALIAÇÃO DO PS-DES UTILIZANDO <i>POOLS</i> HOMOGÊNEOS	68
5.7	ANÁLISE 2: AVALIAÇÃO DO PS-DES UTILIZANDO <i>POOL</i> HETEROGÊNEO	70
5.8	ANÁLISE 3: AVALIAÇÃO DA DIVERSIDADE DE <i>POOLS</i> HOMOGÊNEOS E HETEROGÊNEOS	74
5.9	ANÁLISE 4: AVALIAÇÃO DO PS-DES UTILIZANDO <i>POOL</i> HETEROGÊNEO VS HOMOGÊNEO.	77
5.10	ANÁLISE 5: AVALIAÇÃO DO PS-DES UTILIZANDO DIFERENTES FORMAS PARA CALCULAR O POTENCIAL.	81
5.11	ANÁLISE 6: COMO FUNCIONA PS-DES NA PÓS-SELEÇÃO?	81
5.12	ANÁLISE 7: AVALIAÇÃO DO PS-DES-MVL	84
6	CONCLUSÃO	88
6.1	TRABALHOS FUTUROS	89
6.2	TRABALHOS PUBLICADOS	89
	REFERÊNCIAS	90
	APÊNDICE A – RESULTADOS DA AVALIAÇÃO DO PS-DES UTILIZANDO <i>POOLS</i> HOMOGÊNEOS	96
	APÊNDICE B – ESCOLHAS DAS TÉCNICAS DE SELEÇÃO DINÂMICA DE CLASSIFICADORES PELO PS-DES PARA TODAS AS BASES AVALIADAS	105

1 INTRODUÇÃO

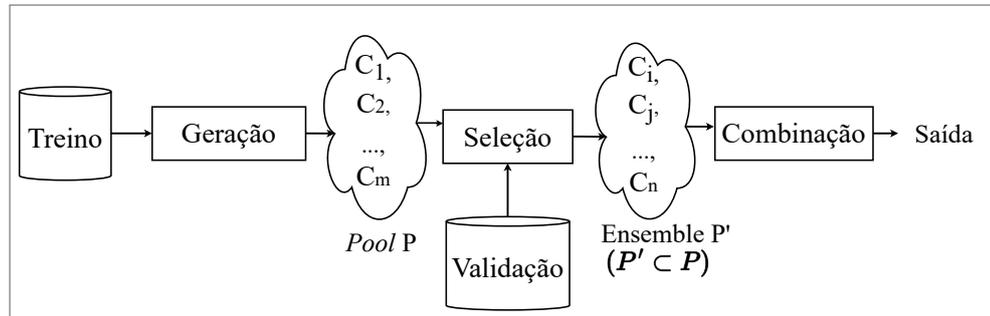
Sistema de múltiplos classificadores (SMC) é uma subárea da Aprendizagem de Máquina, que visa melhorar o desempenho das abordagens que utilizam apenas um classificador, utilizando para isso vários classificadores. (KUNCHEVA, 2014) descreve três razões pelas quais um SMC pode ser superior às abordagens com um único classificador:

- Estatística: Ao optar por utilizar a combinação de múltiplos classificadores é menos provável que se cometam erros na classificação, já que com respostas diferentes para cada classificador, existe uma chance menor de escolher um classificador inadequado.
- Computacional: A utilização de vários classificadores torna menos provável que a busca pela solução fique presa a um mínimo ou máximo local. Isso ocorre porque os classificadores são inicializados de forma distinta e, portanto, exploram melhor o espaço de busca;
- Representacional: A utilização de vários classificadores pode resolver problemas que não seriam possíveis de serem resolvidos em uma abordagem de um classificador. Por exemplo, um classificador linear não é capaz de encontrar a solução ótima para um problema não-linear, contudo, ao combinar diversos classificadores lineares, existe a possibilidade de encontrar uma solução no espaço não-linear.

Tipicamente, um SMC é composto de três fases, como apresentado na Figura 1 (JR; SABOURIN; OLIVEIRA, 2014): (i) geração, no qual é gerado um conjunto de classificadores, comumente denominado de *pool*; (ii) seleção, responsável por selecionar um subconjunto do *pool* gerado na etapa anterior, chamado de *ensemble*; e (iii) combinação ou integração, fase responsável por combinar os classificadores do *ensemble*.

A fase de geração objetiva gerar um *pool* de classificadores-base diversos ou complementares (KUNCHEVA; WHITAKER, 2003), ou seja, classificadores que tenham respostas diferentes para a mesma amostra apresentada. Existem algumas maneiras de atingir a diversidade entre os classificadores no *pool* (DUIN, 2002): treinamento dos modelos em diferentes amostras da base de treinamento como é o caso do *Bootstrap Aggregating*, também conhecido como Bagging (BREIMAN, 1996), diferentes parâmetros para os classificadores-base ou, até mesmo, diferentes modelos de classificação. Um *pool* formado por classificadores-base do mesmo tipo

Figura 1 – As três fases de um Sistema de múltiplos classificadores. Na primeira fase, geração, é formado um *pool* $P = \{C_1, \dots, C_m\}$, com m classificadores. Na segunda fase, Seleção, os classificadores em P são selecionados para formar o *ensemble* $P' = \{C_i, C_j, \dots, C_n\}$, com n , classificadores, de forma que $P' \subset P$.



Fonte: Elaborada pelo autor (2024)

é chamado homogêneo, e, um *pool* formado por classificadores-base de diferentes tipos é denominado heterogêneo (BIAN; WANG, 2007). Outra abordagem que pode auxiliar na geração de *pool* com classificadores diversos é o *Multi-view learning* (MVL) (LI; YANG; ZHANG, 2018b). O MVL analisa diversas visões (*views*) ou representações de dados, obtidas ou geradas, de uma ou de várias bases, visando melhorar o desempenho dos algoritmos de aprendizagem. Existem duas vertentes no campo de MVL: na primeira, os dados já estão dispostos com múltiplas visões. Por exemplo, uma base de dados de faces, contendo imagens RGB e imagens capturadas com câmeras infravermelha, dos mesmos objetos. Assim, diversas técnicas podem ser aplicadas objetivando reduzir as redundâncias entre essas visões, melhorando assim o desempenho dos algoritmos de aprendizagem (XU; TAO; XU, 2013). A segunda vertente, a qual este trabalho se relaciona, trata da construção de múltiplas *views*, geradas a partir da aplicação de funções de representações de dados na base original (SUN et al., 2019b). Essas múltiplas *views* geradas podem ser utilizadas para treinar modelos, aumentando assim, a probabilidade de gerar diversidade entre os classificadores. Porém, a necessidade da aplicação das funções de representação de dados aumenta o custo computacional do sistema.

A fase de seleção de classificadores é dividida em dois grupos (CRUZ; SABOURIN; CAVALCANTI, 2018a): Seleção estática (SEST), na qual o mesmo *ensemble* é definido para todas as amostras de um conjunto de teste, ou Seleção dinâmica (SDIN), na qual é selecionado um *ensemble* para cada amostra em um conjunto de teste. As abordagens SDIN são divididas em duas categorias: Seleção dinâmica de classificador (SDC), quando apenas um classificador é selecionado, e, Seleção dinâmica de *ensemble* (SDE), um dos tópicos centrais desta tese, no qual um *ensemble* é selecionado. A seleção do *ensemble* é guiada utilizando algum

critério, por exemplo: Oráculo (KUNCHEVA, 2002), como é o caso do *K-Nearest-Oracles Eliminate* (KNORA-E) e *K-Nearest-Oracles Union* (KNORA-U) (KO; SABOURIN; JR, 2008), taxa de acerto, como *Dynamic Ensemble Selection Performance* (DES-P) (WOLOSZYNSKI et al., 2012); e meta-aprendizagem (CRUZ et al., 2015), como é o caso do META-DES.

O Oráculo (KUNCHEVA, 2002) é um modelo teórico que escolhe um classificador para cada amostra, se houver pelo menos um classificador no *pool* que possa prever corretamente a classe da amostra. A avaliação do Oráculo é realizada utilizando informações globais, enquanto as técnicas de SDIN devem se concentrar nas informações locais (SOUZA et al., 2017), ou seja, o Oráculo pode trazer informações não relevantes para as técnicas que necessitam de dados específicos sobre uma amostra, obtidos de forma local.

A última fase, combinação, visa integrar as saídas dos classificadores do *ensemble*, selecionado na fase anterior, para produzir uma única predição. Existem diversas técnicas de combinação de classificadores (DUIN, 2002), categorizadas como não-treináveis e treináveis. As técnicas não-treináveis (KITTLER et al., 1998) utilizam regras fixas para combinar os classificadores de um *ensemble*. Entre as técnicas mais utilizadas pode-se citar a média ponderada e o voto majoritário. Já as técnicas treináveis modelam o problema de combinação com um problema de classificação (DUIN, 2002).

É possível encontrar aplicações dos SMC em diversas áreas do conhecimento, por exemplo: financeiro (WANG; ZHANG; YAN, 2023; JUNIOR et al., 2020; XIA et al., 2020; ACHAKZAI; PENG, 2023), saúde (EL-SAPPAGH et al., 2023), engenharia de *software* (JADHAV et al., 2023), agricultura (SWAMINATHAN; PALANI; VAIRAVASUNDARAM, 2022), segurança da informação (BATCHU; SEETHA, 2022; MADDINENI; CH, 2023; RODRIGUES et al., 2023) e interpretação de sinais (VITÓRIO et al., 2023).

1.1 MOTIVAÇÃO

Atualmente as pesquisas na área de Seleção dinâmica concentram-se na proposição de novos métodos para uma das três fases (*i.e.*, geração, seleção ou combinação). Mesmo com o desenvolvimento, tanto teórico quanto aplicado, ainda é possível endereçar alguns pontos de melhoria na área de SDIN, discutidos a seguir. Inicialmente, é importante frisar que existem poucos trabalhos (WANG; CHEN; SUN, 2007; LE; KIM, 2016; KUMAR; MINZ, 2016; WANG et al., 2019; WANG et al., 2021) que exploram a utilização de técnicas de MVL em conjunto com a abordagem de SMC, contudo utilizam apenas estratégias de SEST. Foi encontrado apenas um

trabalho que aplica conceitos de MVL com uma abordagem de SDIN (LU et al., 2022).

Outro ponto de crítica aos trabalhos na área de Seleção dinâmica de *ensemble* é a concentração na proposição de novas formas de selecionar classificadores-base individualmente, *i.e.*, é desenvolvido um novo método que aplicado à fase de seleção para avaliar cada classificador individual no *pool*. Contudo, não existe a preocupação em avaliar o *ensemble* como um todo, ou seja, se os modelos selecionados funcionam bem conjuntamente. Além disso, é possível discutir a falta de esforços científicos em utilizar o conhecimento adquirido nas mais diversas abordagens de seleção dinâmica, ou seja, os diferentes *ensembles* selecionados. Dado que diferentes técnicas de seleção tendem a selecionar diferentes *ensembles*, usar conjuntamente técnicas de seleção pode ser uma caminho viável para melhorar o desempenho dessa categoria de Sistema de múltiplos classificadores. Tal fato se justifica dado que, apesar dos avanços significativos nas pesquisas, não há garantia de que uma única técnica SDE possa obter resultados ótimos na resolução de todos os problemas. Na análise da literatura foi encontrado apenas um trabalho (ELMI; EFTEKHARI, 2021) que propõe a utilização de múltiplas técnicas de SDE, no qual é apresentado o *Multi-Layer Selector* (MLS). Entretanto, o MLS apresenta dois pontos de crítica: (i) não existe a preocupação em avaliar o *ensemble* selecionado em cada camada, pois as técnicas de seleção avaliam um único classificador por vez, (ii) a ordem das técnicas de seleção escolhidas influenciam no resultado final, já que uma técnica SDE pode excluir classificadores competentes, *i.e.*, classificadores que respondem corretamente a uma amostra apresentada, em uma camada.

Diante dos pontos analisados referentes às lacunas nas pesquisas referente às técnicas de Seleção dinâmica de *ensemble* (SDE), esta tese propõe a investigação de duas questões de pesquisa (QP):

- **QP1:** A utilização de um conjunto de técnicas de SDE, especificamente os *ensembles* selecionados pelas mesmas, melhora o desempenho destas técnicas quando utilizadas individualmente?
- **QP2:** A incorporação de técnicas de MVL, individuais e combinadas, especificamente na fase de geração do *pool*, melhor o desempenho de um Sistema de múltiplos classificadores?

Diante das questões de pesquisa abordadas, o objetivo principal desta tese é propor uma abordagem que permita utilizar múltiplas técnicas SDE para aumentar a eficácia da classifica-

ção, quando comparado às técnicas utilizadas individualmente. O objetivo principal pode ser decomposto nos seguintes objetivos específicos:

- Desenvolver uma nova abordagem que avalia e seleciona, dinamicamente, o melhor *ensemble* dado um conjunto de técnicas SDE objetivando gerar classificações mais precisas.
- Desenvolver uma nova forma de avaliação para *ensembles* selecionados por diferentes técnicas de SDE inspirado no Oráculo.
- Avaliar o desempenho da abordagem proposta em diversas bases de dados e realizar a comparação dos resultados com técnicas SDE individuais.
- Avaliar o impacto de diferentes técnicas de geração de *views* individuais e combinadas em um SMC.

1.2 CONTRIBUIÇÕES

As principais contribuições deste trabalho são:

- Uma nova forma de avaliar de *ensembles* denominado *Hardness-aware Oracle* (HaO) que avalia o conjunto de classificadores como um todo, em vez de cada modelo separadamente. O HaO é baseado no Oráculo, contudo, em contraponto ao Oráculo, o HaO é capaz de analisar não apenas a existência de um único classificador competente, mas também a capacidade de uma técnica SDE em encontrar um número significativo desses classificadores. Além disso, o HaO avalia todo o *ensemble* diferentemente das técnicas atuais que avaliam classificador-a-classificador. O HaO foi apresentado em um dos artigos publicados durante a pesquisa (CORDEIRO; CAVALCANTI; CRUZ, 2023b).
- Uma nova abordagem que avalia e seleciona *ensembles* gerados à partir da fase de seleção de múltiplas técnicas SDE, denominada *Post-Selection Dynamic Ensemble Selection* (PS-DES). O PS-DES baseia-se na suposição de que diferentes critérios de seleção podem resultar em diversos *ensembles* selecionados, e os critérios ideais para selecionar um conjunto podem variar dependendo da amostra. O PS-DES opera como uma estratégia de pós-seleção que avalia os *ensembles* selecionados por estas técnicas, utilizando diversas formas de avaliação, como, por exemplo, o HaO, e seleciona a técnicas que obtiveram o melhor resultado na métrica avaliada. Esta contribuição foi apresentada nos

dois trabalhos publicados durante a pesquisa (CORDEIRO; CAVALCANTI; CRUZ, 2023a; CORDEIRO; CAVALCANTI; CRUZ, 2023b).

- Uma nova abordagem que avalia e seleciona *views* geradas a partir de diversas técnicas de transformação de dados, denominado *Post-Selection Dynamic Ensemble Selection with Multi-View Learning* (PS-DES-MVL). O PS-DES-MVL adiciona uma nova camada na geração do *pool* ao PS-DES, para posterior seleção da melhor *view*.

1.3 ESTRUTURA DA TESE

Esta tese conta com a seguinte estrutura de conteúdo: no Capítulo 2 apresenta a fundamentação teórica básica para entendimento do trabalho. O Capítulo 3 apresenta trabalhos publicados nas áreas relacionadas à tese. No Capítulo 4 é apresentado os métodos propostos. O arranjo experimental e os resultados são apresentados no Capítulo 5. A conclusão e os trabalhos futuros estão no Capítulo 6.

2 FUNDAMENTAÇÃO TEÓRICA

Tipicamente, um SMC é composto de três fases: (i) geração, onde é gerado um conjunto de classificadores, comumente denominado de *pool*; (ii) seleção, responsável por selecionar um subconjunto dos classificadores gerados na etapa anterior, comumente chamado de ensemble; e (iii) combinação ou integração, fase responsável por combinar o subconjunto selecionado na fase anterior para finalizar o processo de classificação. As próximas seções detalham cada uma das três principais fases, além de discutir conceitos relacionados, como diversidade, *Multi-view learning*, *Instance Hardness* e Oráculo.

2.1 GERAÇÃO DO *POOL* DE CLASSIFICADORES

A geração do *pool* de classificadores é a fase inicial em um Sistema de múltiplos classificadores. O objetivo é gerar o *pool* P com m classificadores-base C de tal forma que $P = \{c_1, c_2, \dots, c_m\}$. É de fundamental importância que os classificadores-base utilizados nessa fase sejam diferentes entre si, dado que se os classificadores tivessem respostas sempre iguais não seria possível melhorar o desempenho do sistema (ZHOU, 2012). Essa diferença entre os classificadores, presentes no *pool*, é denominada diversidade (ZHOU, 2012), discutido em mais detalhes na Seção 2.4. (DUIN, 2002) elenca seis diferentes formas de gerar classificadores diversos:

- Diferentes inicializações: Essa forma é proposta especificamente para redes neurais, e discute que a configuração iniciais dos pesos pode implicar em diferentes modelos;
- Diferentes parâmetros: Os classificadores-base precisam definir alguns hiperparâmetros, como a quantidade de vizinhos em um *k-Nearest Neighbors* (k-NN), logo, diferentes parâmetros implicam em diferentes modelos gerados.
- Diferentes arquiteturas: Um exemplo dessa forma de gerar classificadores diversos é modificar a quantidade de neurônios na camada escondida em uma *Multilayer Perceptron* (MLP), gerando assim classificadores distintos entre si;
- Diferentes modelos: Este tópico está diretamente relacionado com a tese. A etapa de geração pode utilizar diversos modelos distintos, como por exemplo: k-NN, *Support Vector Machine* (SVM), Árvores de decisão (AD), treinados utilizando a mesma base de

treinamento. Um *pool* formado por classificadores provenientes de diferentes modelos é denominado heterogêneo;

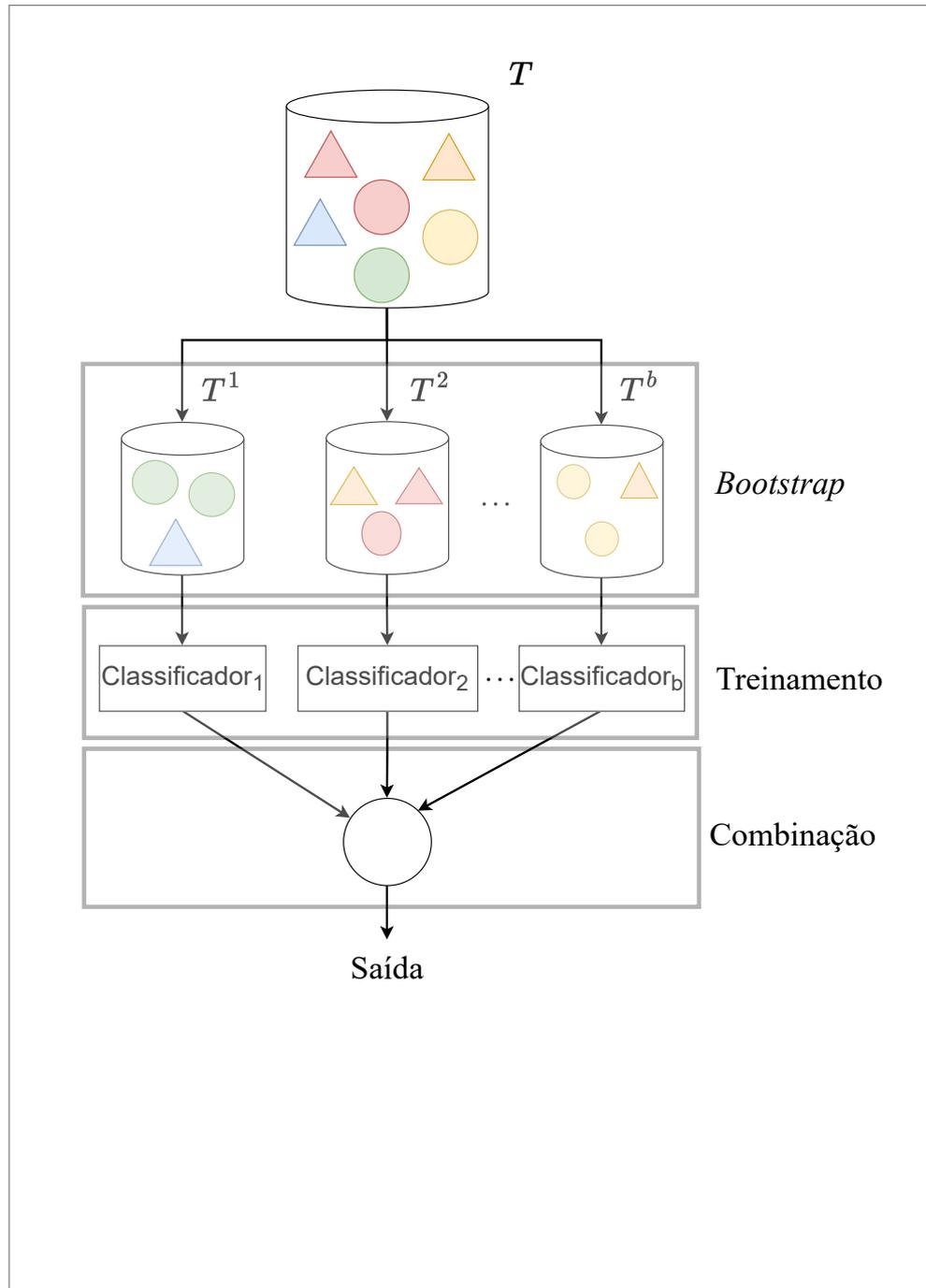
- Diferentes bases de treinamento: Trata da utilização de diferentes amostras, com ou sem reposição, da mesma base de dados, para o treinamento dos modelos de aprendizagem, permitindo que os classificadores se comportem de forma diferente ao serem treinados com diferentes instâncias. Duas técnicas se destacam nesse ponto: Boosting (FREUND; SCHAPIRE, 1997) e o Bagging (BREIMAN, 1996), que será detalhado a seguir.
- Diferentes conjuntos de características: Em algumas situações, uma mesma instância pode ser representada em diferentes espaços de características, como é discutido no *Multi-view learning* (LI; YANG; ZHANG, 2018a). Assim como no item anterior, modelos treinados com diferentes representações, do mesmo objeto, geram diferentes classificadores.

Como abordado anteriormente, uma das técnicas mais utilizadas para gerar diferentes bases de treinamento é o Bagging, que funciona da seguinte forma: dado uma base de treinamento (T), de tamanho N , o objetivo do Bagging é gerar b novas bases T^b , denominados *bootstraps*, adicionando aleatoriamente N instâncias (x) de T , com reposição, o que pode gerar instâncias repetidas em cada uma das novas bases. Depois de gerados, os *bootstraps* são utilizados para treinar b modelos de aprendizagem, aumentando a chance de gerar classificadores com comportamentos diversos. Ao final os classificadores são combinados obtendo uma única saída. Pela proposta de gerar diferentes bases de treinamento, o Bagging é uma boa alternativa para gerar diversidade especialmente quando utilizado em conjunto com classificadores-base instáveis, *i.e.*, classificadores nos quais pequenas alterações nos dados de treinamento influenciam em suas respostas (DIETTERICH, 2000). Exemplos de classificadores instáveis são as Árvores de decisão (AD) e o Perceptron. A Figura 2 mostra o funcionamento do Bagging.

2.1.1 *Multi-view learning*

Multi-view learning (MVL) é uma área que visa entender como diferentes representações de um mesmo dado podem melhorar a eficácia de sistemas de classificação ou predição (LI; YANG; ZHANG, 2018a). Na literatura, é possível encontrar duas abordagens para o MVL. A primeira abordagem tem como pressuposto que o problema a ser estudado é composto por várias bases de dados, obtidas a partir de diferentes fontes contudo derivadas de um mesmo

Figura 2 – Funcionamento do Bagging



Fonte: Elaborada pelo autor (2024)

conjunto de objetos. Por exemplo, em um sistema de segurança pode-se obter informações de impressão digital e da íris de um usuário. O objetivo desta primeira abordagem é encontrar a relação entre as *views*. Na segunda abordagem, denominada construção de *views* (SUN et al., 2019a), o problema é apresentado em uma única base de dados (*i.e.*, em uma única *view*), então são aplicadas diversas técnicas de transformação nos dados, objetivando gerar diversas representações dos dados originais. Essas novas representações devem, preferencialmente, con-

ter características diferentes que podem ser utilizadas para gerar classificações ou previsões mais robustas (ZHAO et al., 2017). Em (SUN et al., 2019a), os autores discutem as principais abordagens para geração das *views*:

- **Partição do conjunto de características:** é uma das abordagens mais simples para a geração de *views*. A proposta é dividir o conjunto de características em diversas partições disjuntas, sem realizar nenhuma transformação nos dados originais. As partições derivadas são consideradas as *views*. Diversas técnicas podem ser aplicadas nesse contexto, entre elas a decomposição aleatória ou algoritmos de otimização (e.g., Algoritmos Genéticos ou Otimização por Enxame de Partículas) para encontrar as melhores partições. Essa abordagem pode reduzir dados redundantes, contudo é importante avaliar o custo computacional da partição das características;
- **Redução de dimensionalidade:** o objetivo dessa abordagem é reduzir a quantidade de dimensões de uma base de dados de forma a evitar dados redundantes ou duplicados e, assim, tentar evitar a maldição da dimensionalidade. Entre as técnicas utilizadas nessa abordagem, pode-se citar o *Principal Component Analysis* (PCA) e o *t-Distributed Stochastic Neighbor Embedding* (t-SNE). Contudo, ao condensar as dimensões, algumas correlações entre as características podem ser perdidas, o que pode impactar negativamente no processo de classificação;
- **Modulação:** também denominada de *Kernel Learning*. Essa abordagem objetiva a aplicação de funções *kernel* para projetar os dados originais em diferentes espaços, gerando assim novas *views*. Entre as possíveis funções *kernel*, pode-se citar a polinomial e a gaussiana. O ponto positivo dessa abordagem é que a transformação gerada nessa abordagem pode melhorar o desempenho dos classificadores. Contudo, existe um custo computacional relacionado à aplicação das funções que deve ser avaliado. Essa abordagem também é discutida em (WANG; CHEN; SUN, 2007), com a aplicação de funções *kernel* linear, de base radial e polinomial.

Além das abordagens mostradas anteriormente, o trabalho apresentado em (PEKALSKA; DUIN, 2001) propõe a geração de *views* à partir de uma representação baseada nas dissimilaridades entre os elementos presentes em uma base de dados. Essas dissimilaridades podem ser calculadas por métricas de distância entre objetos, como a distância Euclidiana ou a distância de Hamming.

2.2 SELEÇÃO DOS CLASSIFICADORES

A segunda fase de um SMC é a seleção. O objetivo dessa fase é selecionar um subconjunto P' , denominado *ensemble*, do pool P , ou seja, $P' \subset P$. Na literatura é possível encontrar duas abordagens (CRUZ; SABOURIN; CAVALCANTI, 2018a): Seleção estática (SEST) e Seleção dinâmica (SDIN). Na abordagem estática, o ensemble P' é selecionado durante a fase de treinamento dado uma base de validação, ou de treinamento, aplicando alguma das estratégias de seleção que serão discutidas a seguir. O mesmo P' será utilizado para classificar todas as amostras de teste. Já na abordagem dinâmica, utilizada nesta tese, o ensemble P' é selecionado especificamente para cada amostra de teste apresentada. Quando apenas um classificador é selecionado denomina-se Seleção dinâmica de classificador (SDC). Quando mais de um classificador é selecionado, denomina-se Seleção dinâmica de *ensemble* (SDE). A Figura 3 mostra as diferenças entre: a) Seleção estática, b) Seleção dinâmica de classificador e c) Seleção dinâmica de *ensemble*. Como a SDE é tema desta dissertação, a próxima subseção é dedicada a aprofundar conceitos relacionados a essa abordagem.

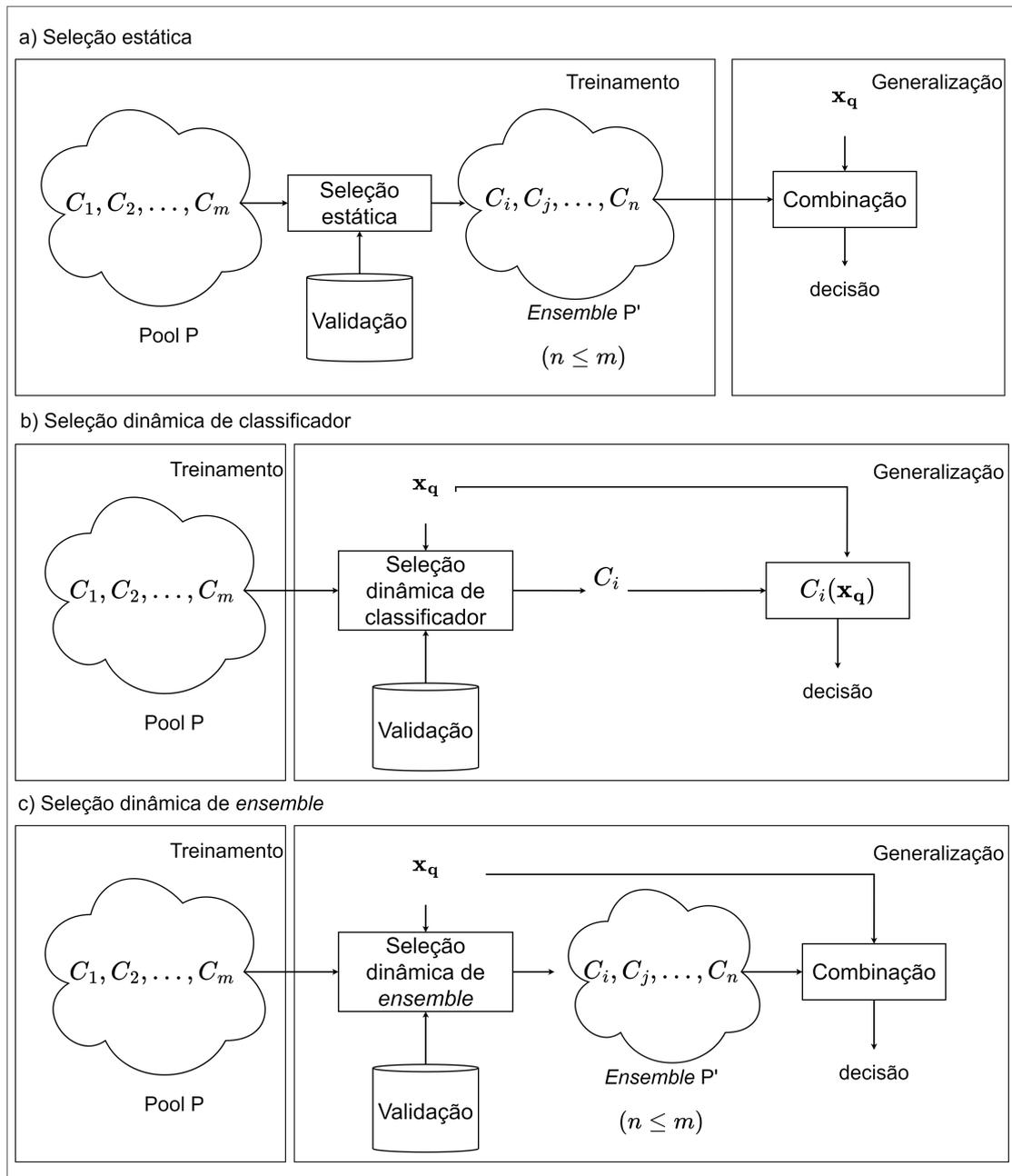
Seleção dinâmica de *ensemble*

A Seleção dinâmica de *ensemble* (SDE) parte do pressuposto que diferentes classificadores podem ser *experts* em diferentes regiões do espaço de busca (CRUZ; SABOURIN; CAVALCANTI, 2018a). Logo, os classificadores são selecionados para cada amostra de teste apresentada. Comumente, as abordagens de SDE funcionam da seguinte forma: primeiro é definido a Região de competência (RdC), então é calculado o grau de competência dos classificadores para a posterior seleção dos classificadores, dado o grau de competência calculado na etapa anterior. Esses dois conceitos são explicados a seguir.

Definição da Região de competência

A Região de competência (RdC) objetiva encontrar padrões similares a uma amostra de teste θ_{x_q} , que estejam em uma base D_{sel} , para, posteriormente, calcular a competência dos classificadores. Para um padrão x_q , a RdC é definida por $\theta_{x_q} = \{x_1, \dots, x_k\}$, onde k é o tamanho da RdC. Entre os métodos para definir a RdC, destacam-se: k-NN (CRUZ et al., 2015), agrupamento (KUNCHEVA, 2000), *Fuzzy Hyperboxes* (DAVTALAB; CRUZ; SABOURIN, 2022) e partição recursiva (SOUZA et al., 2023). O conceito de RdC é um dos mais importantes no contexto da seleção dinâmica de classificadores, já que a sua definição pode interferir na qualidade de abordagens de SDIN (OLIVEIRA; CAVALCANTI; SABOURIN, 2017).

Figura 3 – Diferença entre as abordagens de seleção: a) Seleção estática, b) Seleção dinâmica de classificador e c) Seleção dinâmica de ensemble.



Fonte: Adaptada de (CRUZ; SABOURIN; CAVALCANTI, 2018a)

Definição da competência dos classificadores e seleção

A competência dos classificadores, denominada por $\delta_{C_i, \mathbf{x}_q}$, onde C_i é um classificador base, e \mathbf{x}_q uma amostra avaliada, é um conceito fundamental na área de SDE que objetiva estimar se um classificador será selecionado para fazer parte do *ensemble*. A competência pode ser calculada de diversas formas (e.g., acurácia, *ranking*) a depender do critério de seleção aplicado pela técnica SDE.

2.3 COMBINAÇÃO DOS CLASSIFICADORES

A última fase de um SMC é a combinação, também chamada de agregação ou integração, cujo objetivo é combinar os classificadores selecionados na fase de seleção, quando mais de um classificador é selecionado. Ou seja, sendo o *ensemble* $P' = \{C_i, C_j, \dots, C_n\}$, selecionado na fase anterior, \mathbf{x}_q uma amostra de teste e $\omega(C, \mathbf{x}_q)$ a resposta de um classificador C para \mathbf{x}_q , o objetivo dessa fase é combinar todas as n respostas dos classificadores de P' em uma predição única.

As técnicas de combinação podem ser categorizadas em não-treináveis e treináveis (CRUZ; SABOURIN; CAVALCANTI, 2018a). As técnicas não-treináveis são consideradas mais simples dado que combinam as respostas dos classificadores de acordo com uma regra predefinida. São exemplos de técnicas de combinação não-treináveis (KITTLER et al., 1998; KITTLER, 1998): regra do produto, regra da soma, regra da mediana, voto ponderado e voto majoritário. Entre as regras citadas, o voto majoritário destaca-se por ser uma abordagem comumente utilizada na literatura. O voto majoritário é uma técnica que combina os resultados dos classificadores contando o número de vezes que cada classificador do *ensemble* responde à amostra apresentada, sendo escolhida a classe com maior número de votos. As técnicas não-treináveis podem não ser consideradas ideais para combinação de classificadores, dado que seu desempenho depende do grau de confiança dos classificadores-base (DUIN, 2002).

As técnicas treináveis (ZHOU, 2012) visam aprender a melhor forma de combinar as respostas dos classificadores. Isto significa que, em vez de usar diretamente as saídas dos classificadores-base e aplicar as regras fixas propostas nas técnicas não-treináveis, a abordagem treinável utiliza um modelo de classificação para aprender a melhor forma de combinar essas informações para obter previsões mais precisas (CRUZ; SABOURIN; CAVALCANTI, 2018a). Dado

que é necessário realizar mais um processo de aprendizagem, as técnicas treináveis tendem a ter maior custo computacional quando comparadas a técnicas não-treináveis. Entre exemplos de técnicas treináveis, pode-se citar o *Stacking* (ZHOU, 2012) e *Mixture of experts* (MASOUDNIA; EBRAHIMPOUR, 2014).

2.4 DIVERSIDADE

Como discutido anteriormente, em um SMC é interessante que os classificadores treinados na fase de geração sejam diferentes entre si, ou seja, obtenham respostas distintas para a mesma amostra. Essa diferença entre os classificadores é denominada diversidade, uma questão relevante no âmbito dos SMC (KUNCHEVA, 2014). Não existe nenhuma definição formal sobre diversidade, contudo, foram desenvolvidas diversas métricas que contribuem para entender melhor essa característica (KUNCHEVA; WHITAKER, 2003). As métricas de diversidade são divididas em duas categorias: pareada, calculada entre dois classificadores, e não-pareada, que considera todos os classificadores de uma única vez. As próximas subseções abordarão detalhes de técnicas pareadas e técnicas não-pareadas.

2.4.1 Métricas pareadas

As medidas pareadas calculam a diversidade para um par de classificadores. Um *pool* com m classificadores produz $m(m - 1)/2$ valores de diversidade. Para calcular um valor único de diversidade para o *pool* é necessário calcular o valor médio entre todos os valores individuais, tornando uma métrica global. As métricas pareadas, geralmente, calculam a diversidade dado as relações entre pares de classificador em situações onde classificam corretamente (1) ou erroneamente (0) uma amostra apresentada. A Tabela 1 resume as possíveis relações entre pares de classificadores.

Tabela 1 – Relação entre pares de classificadores

	C_k correto (1)	C_k errado (0)
C_i correto (1)	N^{11}	N^{10}
C_i errado (0)	N^{01}	N^{00}

Fonte: Elaborada pelo autor (2024)

Entre as métricas pareadas pode-se citar (KUNCHEVA; WHITAKER, 2003):

Métrica *Disagreement* (DIS): Essa métrica de diversidade calcula a razão entre o número de vezes que um classificador (C_k) acerta uma amostra, e o outro classificador (C_j) erra a mesma amostra, pelo total de amostras, representado pela Equação 2.1:

$$DIS_{C_k, C_j} = \frac{N^{10} + N^{01}}{N^{11} + N^{10} + N^{01} + N^{00}} \quad (2.1)$$

Quanto maior o valor de DIS, maior é a diversidade entre os par de classificadores avaliado.

Métrica *Double-Fault* (DF): Essa métrica visa avaliar a diversidade a partir da razão entre a quantidade de vezes que os dois classificadores erram uma amostra e a quantidade total de amostras, representado pela Equação 2.2:

$$DF_{C_k, C_j} = \frac{N^{00}}{N^{11} + N^{10} + N^{01} + N^{00}} \quad (2.2)$$

Quanto maior o valor de DF, menor é a diversidade entre o par de classificadores avaliado.

É possível calcular a Diversidade global (DG) à partir de uma métrica pareada. Para tal, é necessário calcular a média da métrica pareada para todos os pares de classificadores. Sendo m é o tamanho do *pool*, e Q_{C_i, C_k} uma métrica pareada calculada para dois classificadores C_i e C_k , a diversidade global é calculada utilizando a Equação 2.3:

$$DG = \frac{2}{m(m-1)} \sum_{i=1}^{m-1} \sum_{k=i+1}^m Q_{C_i, C_k} \quad (2.3)$$

2.4.2 Métricas não-pareadas

Esse tipo de métrica visa calcular um único valor de diversidade para todo o *pool*. Entre as diversas métricas propostas na literatura (KUNCHEVA, 2014), citam-se duas:

Métrica *Entropy* (ET): Se todos os classificadores classificarem da mesma forma um padrão apresentado não existe diversidade. Partindo desse pressuposto, dado uma amostra \mathbf{x}_i , sendo m o tamanho do *pool*, e n a quantidade de amostras em uma base, e $y_{i,j}$ a resposta de um classificador C_i a uma amostra \mathbf{x}_i , a métrica ET é calculada de acordo com a Equação 2.4:

$$ET = \frac{1}{N} \frac{2}{m-1} \sum_{j=1}^n \min \left\{ \left(\sum_{i=1}^m \right), \left(m - \sum_{i=1}^m \right) \right\} \quad (2.4)$$

Os valores de ET variam de 0, que significa que não existe diversidade, a 1, que significa a maior diversidade possível.

Métrica Kohavi–Wolpert Variance (KWV): Essa métrica não-pareada de diversidade é derivada a partir da média da métrica (*Disagreement*). Sendo m o tamanho do *pool* e DIS_{med} o valor médio da métrica DIS, KWV é calculada pela Equação:

$$KV = \frac{m - 1}{2m} \times DIS_{med} \quad (2.5)$$

Quanto maior o valor de KWV, maior é a diversidade entre o par de classificadores avaliado.

2.5 ORÁCULO

O Oráculo (KUNCHEVA, 2002) foi introduzido como um modelo conceitual que escolhe um classificador capaz de definir uma resposta correta à uma amostra, dada a presença de pelo menos um classificador que a classifica corretamente no *pool*. Por ser o modelo de seleção perfeito, o Oráculo é considerado um possível limite superior para as técnicas de SDC.

Embora abstrato, o Oráculo encontra aplicação prática em várias fases de um SMC. Por exemplo, o trabalho (SOUZA et al., 2017) propõe aplicar o conceito do Oráculo na fase de geração, gerando hiperplanos iterativamente à partir de Perceptrons, para garantir que cada instância no conjunto de treinamento seja classificada com precisão por pelo menos um dos classificadores-base do *pool*. Os autores concluíram que o Oráculo tem acesso a informações globais contudo as técnicas SDIN tratam com informações locais, referentes a cada instância, o que dificulta a obtenção de um desempenho semelhante ao do Oráculo. O Oráculo também é aplicado na fase de seleção, exemplificado pelo KNORA-U e KNORA-E (KO; SABOURIN; JR, 2008), discutidos anteriormente.

2.6 INSTANCE HARDNESS

O *Instance Hardness* (IH) (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014) é uma característica de uma amostra que calcula probabilidade dessa amostra ser classificada erroneamente por um classificador. Por exemplo, amostras consideradas *outliers*, ou que têm sua classe definida de forma errada, comumente, são mais difíceis de serem classificadas corretamente, ou seja, possuem um maior valor de IH. Foram propostas diversas medidas para IH, como por exemplo: *k-Disagreeing Neighbors* (kDN), *Disjunct Class Percentage* (DCP) e *Class Likelihood* (CL).

Como o kDN mede a sobreposição local de uma amostra em relação aos seus vizinhos mais próximos, ou seja, reflete a dificuldade de classificação decorrente da sobreposição local de classes na região da amostra, é comumente aplicada nas abordagens de SDIN. O kDN é calculado seguindo a Equação 2.6, onde \mathbf{x}_i é a amostra, T é a base de dados, e k é a quantidade de vizinhos mais próximos:

$$kDN(\mathbf{x}, T, k) = \frac{|\mathbf{x}_j \in kNN(\mathbf{x}_i, T, k_h) \wedge label(\mathbf{x}_j) \neq label(\mathbf{x}_i)|}{k_h} \quad (2.6)$$

Logo, o kDN é uma métrica que varia de 0 a 1. Quanto menor o valor do kDN, maior a probabilidade de um classificador responder corretamente a amostra apresentada.

Alguns trabalhos aplicam conceitos IH aos SMC, tanto na fase de geração dos classificadores como na seleção. (WALMSLEY et al., 2018) propôs o Bagging-IH que utiliza conceitos de IH na fase de geração dos classificadores, em conjunto com o Bagging, de forma que a seleção de uma instância durante o processo de re-amostragem é inversamente proporcional ao seu IH. O sBal_IH (CHONGOMWERU; KASEM, 2021) também utiliza conceitos de IH em conjunto com o Bagging, contudo o foco é na geração de *bootstraps* mais balanceadas para problemas com bases de dados desbalanceadas. (SOUZA et al., 2022) aplica conceitos de IH para a fase de seleção baseado no KNORA-E. A proposta é utilizar medidas de IH para remover instâncias consideradas mais difíceis da RdC, melhorando assim a seleção dos classificadores.

3 TRABALHOS RELACIONADOS

Este capítulo apresenta alguns trabalhos relacionados a área de Sistema de múltiplos classificadores. São apresentados trabalhos relacionados aos SMC que utilizam *pools* de classificadores homogêneos e heterogêneos que introduziram abordagens inovadoras. Também são listados alguns trabalhos da área de MVL em interseção com SMC. Ao final, é proposta são discutidos os tópicos abordados neste capítulo.

3.1 TRABALHOS COM CLASSIFICADORES HOMOGÊNEOS

Nesta seção são discutidos trabalhos relacionados ao Sistema de múltiplos classificadores (SMC), que utilizam classificadores-base homogêneos. A Tabela 2 apresenta um resumo com a técnica desenvolvida, ano de publicação, classificador-base utilizado e tipo de seleção. Os artigos apresentados na Tabela 2 e não apresentados nessa seção, serão discutidos na próxima seção pois também utilizam classificadores heterogêneos.

From dynamic classifier selection to dynamic ensemble selection (KO; SABOURIN; JR, 2008). Este trabalho apresenta o *K-Nearest-Oracles Eliminate* (KNORA-E) uma das técnicas comumente citadas na literatura referente à SDE. Essa técnica seleciona o *ensemble* à partir do conceito do Oráculo, de forma que, dada uma Região de competência, serão selecionados apenas os classificadores que consigam classificar corretamente todas as amostras nesta RdC. Caso não seja possível encontrar nenhum classificador que corresponda ao que foi definido, o tamanho da região de competência é diminuído. Caso mais de um classificador seja escolhido, a combinação é realizada utilizando o voto majoritário. Também é apresentado o *K-Nearest-Oracles Union* (KNORA-U), que segue o conceito do Oráculo, como discutido para o KNORA-E. A diferença encontra-se na forma de selecionar: dada uma RdC, o KNORA-U seleciona os classificadores que consigam classificar corretamente pelo menos uma amostra na RdC. Com isso, é possível que um classificador seja selecionado mais de uma vez, fazendo com que tenha maior peso no momento da combinação. Então, utilizando o voto majoritário como abordagem de combinação, os classificadores são combinados.

A measure of competence based on random classification for dynamic ensemble selection (WOLOSZYNSKI et al., 2012). Este artigo apresenta o *Dynamic Ensemble Selection Performance* (DES-P), uma técnica SDE *ensemble* que utiliza o conceito de classificador alea-

tório: uma abordagem que aleatoriamente escolhe uma classe para uma amostra apresentada. Dado uma RdC, a competência de cada classificador no *pool* é calculada como a diferença entre a acurácia do classificador subtraída da acurácia de um classificador aleatório. Apenas os classificadores que se comportam melhores que o classificador aleatório são selecionado para fazer parte do *ensemble*.

Dynamic selection approaches for multiple classifier systems (CAVALIN; SABOURIN; SUEN, 2013). Este trabalho apresenta o *K-Nearest Output Profile* (KNOP), uma abordagem que visa selecionar todos os classificadores que classificam corretamente pelo menos uma amostra pertencente à RdC. Para isso, a RdC é estimada utilizando as decisões do classificador-base, ou seja, as classes previstas. Dessa forma, a similaridade entre a amostra avaliada e as amostras da RdC é medida no espaço de decisão, em vez do espaço de características.

Meta-DES: A dynamic ensemble selection framework using meta-learning (CRUZ et al., 2015). O META-DES é uma abordagem SDE que utiliza o conceito de *meta-learning* para selecionar os classificadores, utilizando para tal um meta-classificador. O META-DES funciona em três fases: geração dos classificadores, meta-treinamento e generalização. Na geração é utilizado o Bagging para criação do *pool* de classificadores. Na fase do meta-treinamento, o objetivo é treinar um meta-classificador (MLP) que definirá os classificadores que serão selecionados na próxima fase. O processo de treinamento do meta-classificador consiste em utilizar cinco meta-características dos classificadores-base, calculadas para cada amostra em um base de treinamento, vinculadas a saída do classificador. Na última fase, generalização, o objetivo é definir a classe de uma amostra de teste apresentada. Para tal, primeiro é realizada a seleção de classificadores, utilizando o meta-classificador, onde cada classificador é avaliado para definir se fará parte do *ensemble*. No final, os classificadores selecionados são combinados utilizando o voto majoritário.

Margin & diversity based ordering ensemble pruning (GUO et al., 2018). Este trabalho propõe utilizar o conceito de teoria das margens e diversidade como forma de ranquear os classificadores para então selecionar, de forma estática, os melhores classificadores para um *ensemble*. A técnica proposta no artigo é o *Margin & Diversity based ordering Ensemble Pruning* (MDEP). A quantidade de classificadores selecionados é um parâmetro da proposta. O *pool* é formado por 200 AD, gerados utilizando Bagging. A combinação é feita pelo voto majoritário. A proposta é avaliada com Bagging e outras técnicas de ranqueamento.

A weighted multiple classifier framework based on random projection (NGUYEN et al., 2019). O trabalho propõe uma abordagem de Seleção estática, que utiliza a proje-

ção aleatório como ferramenta para gerar um *pool* de classificadores-base (AD), denominada *Weighted Multi-classifiers Framework Based on Random Projection* (WMCRP). Não existe seleção de classificadores, contudo, a combinação utilizando uma abordagem ponderada permite que classificadores menos competentes tenham pouca participação no processo decisório.

Ensemble selection with joint spectral clustering and structural sparsity (WANG et al., 2021). Os autores utilizam de conceitos de agrupamento espectral para gerar esparsidade estrutural para a seleção estática de *ensemble*, utilizando aprendizado não-supervisionado, com uma técnica denominada *Ensemble Selection with Joint Spectral Clustering and Structural Sparsity* (ES-JSS). Um dos pontos discutidos no trabalho é que a proposta é menos sensível a mudanças nos tipos de classificadores-base utilizados, concluindo que a principal informação para a escolha de *ensembles* estáticos está no espaço de predição dos modelos individuais.

Multi-Layer Selector(MLS): Dynamic selection based on filtering some competence measures (ELMI; EFTEKHARI, 2021). O MLS é uma proposta de combinar métodos de SDIN em diversas camadas para melhorar o processo de seleção do *ensemble*. O MLS aplica o método *Bootstrap* na fase de geração, como no Bagging, gerando um *pool* com 100 Perceptrons. Na fase de seleção, o MLS funciona em várias camadas que aplicam SDE para selecionar classificadores, passando-os para a camada subsequente. Os métodos SDE são integrados à estrutura do MLS, de forma a colaborar para excluir classificadores que não sejam competentes em cada camada. Na fase de combinação, é aplicado o voto majoritário para integrar os classificadores selecionados.

A scalable dynamic ensemble selection using fuzzy hyperboxes (DAVTALAB; CRUZ; SABOURIN, 2024). O *Fuzzy Hyperbox Dynamic Ensemble Selection* (FH-DES) é uma abordagem que utiliza conceitos de *fuzzy hyperboxes* para calcular a RdC, em contraponto ao k-NN que é mais comumente utilizado. Os resultados dos experimentos, configurados um *pool* homogêneo com 100 classificadores-base Perceptron, mostram que o FH-DES obteve resultados superiores aos algoritmos tradicionais de SDE além de diminuir o custo computacional. Uma atualização na proposta também é apresentada objetivando melhorar a capacidade dos métodos SDE para grandes conjuntos de dados, incorporando um processo de contração juntamente com o mecanismo de expansão existente para o cálculo da RdC.

Tabela 2 – Resumo das informações (ano de publicação, classificadores base e métodos de seleção) relacionados as técnicas de SMC utilizando classificadores-base homogêneos.

Técnica	Ano	Classificador	Seleção
KNORA-E (KO; SABOURIN; JR, 2008)	2008	k-NN, Parzen, QDC	SDE
KNORA-U (KO; SABOURIN; JR, 2008)	2008	k-NN, Parzen, QDC	SDE
DES-CV (WOLOSZYNSKI; KURZYNSKI, 2011)	2011	AD	SDE
DES-P (WOLOSZYNSKI et al., 2012)	2012	AD	SDE
KNOP (CAVALIN; SABOURIN; SUEN, 2013)	2013	AD, k-NN	SDE
META-DES (CRUZ et al., 2015)	2015	Perceptron	SDE
MDEP (GUO et al., 2018)	2018	AD	SEST
WMCRP (NGUYEN et al., 2019)	2019	AD	SEST
ES-JSS (WANG et al., 2021)	2021	AD	SEST
MLS (ELMI; EFTEKHARI, 2021)	2021	Perceptron	SDE
FH-DES (DAVTALAB; CRUZ; SABOURIN, 2024)	2024	Perceptron	SDE

Fonte: Elaborada pelo autor (2024)

3.2 TRABALHOS DE SISTEMA DE MÚLTIPLOS CLASSIFICADORES COM CLASSIFICADORES HETEROGÊNEOS

Nesta seção são discutidos trabalhos relacionados aos Sistema de múltiplos classificadores (SMC), que utilizam classificadores-base heterogêneos. A Tabela 4 apresenta um resumo com o nome da técnica desenvolvida no trabalho, ano de publicação, classificadores-base utilizados e tipo de seleção. As siglas, que ainda não foram citadas, referentes aos classificadores-base encontram-se na Tabela 3.

Selective fusion of heterogeneous classifiers (TSOUMAKAS; ANGELIS; VLAHAVAS, 2005). Este artigo propõe uma nova forma de seleção para um *pool* formado por classificadores heterogêneos, denominada *Selective Fusion* (SF). A seleção envolve a geração de estimativas independentes de erro para cada classificador, a aplicação de testes estatísticos de múltiplas comparações (Tukey, Hsu e Scott and Knott) para avaliar essas estimativas e a identificação do conjunto de classificadores com a menor taxa de erro média. A abordagem foi aplicada a um *pool* heterogêneo gerado com 10 diferentes modelos de aprendizagem, os quais não fizeram nenhum tipo de ajuste de parâmetros, que posteriormente são combinados utilizando votação ponderada.

A probabilistic model of classifier competence for dynamic ensemble selection (WOLOSZYNSKI; KURZYNSKI, 2011). Este trabalho propõe uma nova maneira de calcular a Região

Tabela 3 – Lista de classificadores, e suas respectivas siglas, utilizadas nos trabalhos que utilizam *ensembles* heterogêneos.

Classificador	Acrônimo
<i>Adacost</i>	ADAC
Bagging	BAG
<i>Cost-Sensitive CART</i>	C-CART
<i>Cost-Sensitive Random Forest</i>	C-RF
<i>Decision Table</i>	DET
<i>Decision Stump</i>	DS
<i>Deep Neural Network</i>	DNN
<i>Fisher Classifier</i>	FC
<i>Gaussian Process Classifier</i>	GPC
<i>Gaussian Bayes</i>	GB
<i>Gaussian Process Classifier</i>	GPC
<i>KStar</i>	k*
<i>KStar</i>	LDA
<i>Linear Discriminant Classifier</i>	LDC
<i>Logistic Classifier</i>	LC
<i>Logistic Regression</i>	LR
<i>Nearest Mean Classifier</i>	NMC
<i>PART Rule Learning Algorithm</i>	PART
<i>Quadratic Discriminant Analysis</i>	QDA
<i>Quadratic Discriminant Classifiers</i>	QDC
<i>Radial Basis Function Network</i>	RBF
<i>Random Subspace Method</i>	RSM
<i>Rotation Forest</i>	RTF
<i>Stochastic Gradient Boosting</i>	SGB
<i>RIPPER rule learning</i>	JRIP
<i>Sequential Minimal Optimization</i>	SMO
<i>Support Vector Machine with Quadratic Kernel</i>	SVMQ
<i>Trimmed Bagging</i>	TBAG

Fonte: Elaborada pelo autor (2024)

de competência, denominada *DES-competence based on label outputs and weighted majority voting* (DES-CV), no contexto de seleção dinâmica, utilizando uma função baseada em conceitos probabilísticos, e conseqüentemente, selecionar e combinar, ponderando a partir das competências. A abordagem avalia o caso da Seleção dinâmica de classificador, assim como a Seleção dinâmica de *ensemble*. Os autores realizaram experimentos comparando a abordagem proposta com técnicas de seleção dinâmica, como é o caso do DCS-LA e KNORA-E. Além da análise da RdC, também foi avaliado o comportamento quando utilizado *pools* formados por classificadores homogêneos (AD) e heterogêneos (k-NN, SVM, NMM e Parzen). Não foram realizadas comparações entre os *pools* heterogêneos e homogêneos. Contudo, os resultados apresentados mostram que a abordagem proposta, utilizando a SDE, supera alguns dos algoritmos mais consolidados na área.

Toward a General-Purpose Heterogeneous Ensemble for Pattern Classification (NANNI et al., 2015). Este trabalho não propõe nenhuma novidade em relação a nenhuma das fases de um SMC. Contudo, ele realiza um estudo em busca do *General-Purpose* (GP), comparando o impacto da eficácia dos *ensembles* heterogêneos, compostos por diferentes combinação de classificadores, sem aplicação da seleção, combinados com a regra da soma. Os autores concluem que a combinação de diferentes modelos de classificadores é a maneira mais viável de criar um *pool* mais generalista.

Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification (HAQUE et al., 2016). A inovação proposta neste trabalho é a utilização de um algoritmo genético para encontrar o melhor subconjunto dado um *pool* heterogêneo gerado com 20 classificadores-base, utilizado para solucionar problemas em base de dados desbalanceadas. A abordagem é denominada *GA-based searching for ensemble combination* (GA-EoC). O objetivo do Algoritmo Genético é selecionar um *ensemble* utilizando o *Matthews Correlation Coefficient* (MCC) como função a ser otimizada. O voto majoritário é utilizado para integrar os classificadores. De acordo com os experimentos apresentados no artigo, a abordagem proposta obteve melhores resultados dos que os classificadores, analisados individualmente, além de obter resultados superiores a métodos de múltiplos classificadores como o Bagging e *Random Forest* (RF). Os autores mostraram que para diferentes bases de dados, a abordagem consegue selecionar diferentes classificadores-base.

The Heterogeneous Ensembles of Standard Classification Algorithms (HESCA): the Whole is Greater than the Sum of its Parts (LARGE; LINES; BAGNALL, 2017). Este artigo avalia o impacto de dois *pools* formados por classificadores heterogêneos, um com

cinco classificadores-base, o outro com nove classificadores-base. Não é aplicada nenhuma abordagem de seleção, apenas a combinação dos classificadores utilizando uma abordagem ponderada a partir das probabilidades dos classificadores-base. O artigo não propõe nenhuma nova forma de gerar, selecionar ou combinar, contudo, discute o desempenho superior da utilização de *pools* heterogêneos, quando comparados a abordagens homogêneas.

Combining multiple algorithms in classifier ensembles using generalized mixture functions (COSTA et al., 2018). Os autores propuseram a utilização de funções de mistura generalizadas como uma nova forma de calcular pesos para combinação de classificadores dinamicamente, sem a necessidade de realizar a seleção prévia dos classificadores. Na fase de geração do *pool* de classificadores, foram utilizados cinco classificadores-base distintos, além da utilização do Bagging. Os experimentos, que avaliam apenas bases binárias, mostraram que a abordagem proposta obtém resultados superiores quando comparadas com diversos SMC sedimentados na literatura (e.g., RF, META-DES)

Combining heterogeneous classifiers via granular prototypes (NGUYEN et al., 2018). O trabalho utiliza conceitos de granularidade, ou seja, a interpretação de dados complexos em intervalos numéricos. A granularidade é aplicada a cada classe do problema, combinado com lógica *fuzzy*, a partir da união de todos os classificadores do *pool*. Posteriormente é calculada a distância entre os padrões e os meta-dados para a base de treinamento. Os autores propuseram aplicar esses conceitos em duas configurações de *pool*: uma composta por 3 classificadores-base e a outra com 10 classificadores-base. Os resultados mostram a superioridade da abordagem proposta quando comparada a outras técnicas como Adaboost e *Random Forest*. Seria interessante comparar o que foi proposto com outras técnicas como o KNORA-U ou META-DES. O artigo discute a dificuldade em definir a quantidade ótima de classificadores-base presentes no *pool*, já que, quando compara-se as configurações com 3 e 10 classificadores, nenhuma consegue obter os melhores resultados em todas as bases.

Investigating the impact of selection criteria in dynamic ensemble selection methods (FILHO; CANUTO; SANTIAGO, 2018). Este trabalho investiga o impacto de critérios para seleção dinâmica de classificadores. São propostas três abordagens, aplicadas após a seleção dos algoritmos KNORA-U e KNORA-E. Essas abordagens realizam a: (i) variação da quantidade de classificadores selecionados, utilizando um parâmetro percentual, denominado *KNORA-PERCENTAGE* (KP) (ii) seleção aleatória, e, (iii) seleção utilizando conceitos de distância, denominada *KNORA-distance dynamic ensemble selection* (KNP-DES). Foram realizados experimentos em 16 cenários diferentes, modificando os parâmetros das três abor-

dagens citadas anteriormente, comparados a diversas técnicas SDE, como o META-DES e RF, no qual a abordagem proposta, utilizando a seleção com conceitos de distância com a métrica de diversidade *Double-Fault*, obteve os melhores resultados. Contudo, a avaliação foi baseada apenas na taxa de acerto, necessitando outras métricas avaliativas para conclusão da qualidade da abordagem proposta.

Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach (BOCK; COUSSEMENT; LESSMANN, 2020). A utilização de *ensembles* heterogêneos para previsão de problemas financeiros em negócios é abordada neste trabalho, com uma técnica denominada *Cost-Sensitive Multicriteria Ensemble Selection* (CSMES). Com um *pool* composto por 20 classificadores-base, os autores abordam o problema de selecionar os classificadores utilizando uma abordagem de otimização multi-objetivo, maximizando a taxa de verdadeiro positivo e minimizando a taxa de falso negativo. Os classificadores selecionados são combinados pelo método de média. Apesar de apresentar a comparação entre diversas técnicas de otimização multi-objetivo, o artigo não discute se a abordagem proposta é superior quando comparada aos métodos tradicionais de SMC.

HBoost: A heterogeneous ensemble classifier based on the Boosting and entropy measurement (KADKHODAEI; MOGHADAM; DEGHAN, 2020). Este trabalho propõe o *Heterogeneous Boosting-based ensemble classifier* (HBoost) que aplica a seleção dos classificadores em um *pool* de 100 classificadores, gerados à partir da aplicação do Boosting em 10 classificadores-base distintos. A partir da ordenação dos classificadores por acurácia, a proposta seleciona um *ensemble* com alta taxa de acurácia e alta diversidade, tornando a tarefa um problema combinatório. Este é o grande gargalo da abordagem, dado que é realizada a busca entre todas as possíveis combinações de classificadores selecionados. Contudo, como é um problema estático, essa busca é realizada apenas uma vez. A abordagem obteve resultados melhores, na maior parte das bases, quando comparados a abordagens como um classificador, assim como a abordagens de múltiplos classificadores como Bagging, Adaboost e Stacking.

HDEC: A heterogeneous dynamic ensemble classifier for binary datasets (OSTVAR; MOGHADAM, 2020). Este trabalho propõe o *Heterogeneous Dynamic Ensemble Classifier* (HDEC) uma abordagem de seleção dinâmica de classificadores para bases de dados binárias. O HDEC as taxas de verdadeiro positivo e verdadeiro negativo para separar os classificadores, formando assim dois *ensembles* de classificadores: os que classificam melhor exemplos positivos, e os que classificam melhor os exemplos negativos, baseados na base de validação. Os

classificadores selecionados são combinados pela regra da soma. Os experimentos, mesmo que sendo avaliadas poucas bases de dados, mostram que o HDEC é superior a técnicas como o Bagging e o Adaboost, contudo, as técnicas comparadas utilizavam *pools* homogêneos.

Preprocessed dynamic classifier ensemble selection for highly imbalanced drifted data streams (ZYBLEWSKI; SABOURIN; WOŹNIAK, 2021). A novidade deste trabalho é a aplicação do SDE a fluxos de dados desbalanceados. Não existe nenhuma melhoria nas fases de geração, seleção ou combinação. Na geração, são utilizados quatro classificadores-base, acrescidos do Bagging, selecionados por duas técnicas SDE: KNORA-U e KNORA-E. Os classificadores selecionados são combinados com a regra da soma.

A hierarchical fusion framework to integrate homogeneous and heterogeneous classifiers for medical decision-making (WANG et al., 2021). Este trabalho propõe o Hierarchical fusion of homogeneous and heterogeneous multi-classifiers (HF2HM), uma abordagem focada na combinação de *pool* de classificadores. São discutidos dois níveis: no primeiro nível são gerados diversos *pools* homogêneos utilizando, para cada um dos sete classificadores-base, representações dos dados de treinamento a partir de *random projection matrices*. No segundo nível é realizada a combinação, de forma dinâmica, dos diversos *pools* homogêneos pela matriz de probabilidade à posteriori. A abordagem é comparada com cenários com um único classificador, além de abordagens de múltiplos classificadores, como o Bagging e Adaboost, obtendo resultados superiores em diversos tipo de base de dados.

Dynamic ensemble learning for multi-label classification (ZHU et al., 2023). Este trabalho propõe uma nova abordagem SDE para problemas multi-classes, *i.e.*, problemas que uma amostra pode pertencer a mais de uma classe. A abordagem é denominada *Multi-Label classification with Dynamic Ensemble learning* (MLDE). A geração dos classificadores utiliza o Bagging para gerar diversidade no *pool*, aplicado a diferentes bases, divididas pela quantidade de classes, dada a base de treinamento original. A novidade do trabalho é a forma de calcular a competência dos classificadores que serão selecionados para o *ensemble* utilizando a combinação da acurácia e do *ranking loss*, uma forma de avaliar a correlação entre as diferentes classes. Os experimentos são realizados com um *pool* heterogêneo, e a abordagem proposta consegue obter resultados melhores, em seis métricas avaliadas, do que outras técnicas multi-classe.

Tabela 4 – Resumo das informações (ano de publicação, classificadores base e métodos de seleção) relacionados as técnicas de SMC utilizando classificadores-base heterogêneos. As técnicas que não tem nomes definidos no trabalho estão como S/N.

Técnica	Ano	Classificadores	Seleção
SF (TSOUMAKAS; ANGELIS; VLAHAVAS, 2005)	2005	DET, JRIP, PART, AD, k-NN, k*, NB, SMO, RBF, MLP	SEST
DES-CV (WOLOSZYNSKI; KURZYNSKI, 2011)	2011	LC, NMC, k-NN, Parzen Classifier, AD, SVM	SDE
S/N (NANNI et al., 2015)	2015	SVM, GPC, AdaBoost	Não
GA-EoC (HAQUE et al., 2016)	2016	NB, SVM, LC, k-NN, DET, AD, RF	Não
HESCA (LARGE; LINES; BAGNALL, 2017)	2017	LC, C4.5, SVM, k-NN, MLP, RF, RTF, SVMQ, DNN	Não
S/N (COSTA et al., 2018)	2018	k-NN, MLP, AD, NB, SVM	Não
S/N (NGUYEN et al., 2018)	2018	LDA, NB, k-NN, C.45, DS, FC, NMC, LC	Não
S/N (FILHO; CANUTO; SANTIAGO, 2018)	2018	MLP, NB, AD, k-NN, SVM	SDE
CSMES (BOCK; COUSSEMENT; LESSMANN, 2020)	2020	BAG, TBAG, SGB, RTF, RSM, RF, AD, C4.5, C4.4, LR, LDA, QDA, MLP, SVM, k-NN, ADAC, C4.5 + MC, C-RF, C-AD	SEST
HBoost (KADKHODAEI; MOGHADAM; DEGHAN, 2020)	2020	NB, BN, k*, k-NN, AD, RT, DS, 1R, DET, SVM	SEST
HDEC(OSTVAR; MOGHADAM, 2020)	2020	NB, BN, k*, k-NN, AD, RT, DS, 1R, DET, SVM	SDE
S/N (ZYBLEWSKI; SABOURIN; WOŹNIAK, 2021)	2021	GB, HT, k-NN, SVM	SDE
HF2HM (WANG et al., 2021)	2021	GB, LR, QDA, k-NN, AD, RF, XGBoost	Não
MLDE (ZHU et al., 2023)	2023	Adaboost, AD, Bernoulli NB, GB, extra trees, multinomial NB, linear SVM, kernel SVM, RF, XGBoost, SGB, LDA, QDA	SDE

Fonte: Elaborada pelo autor (2024)

3.3 TRABALHOS SOBRE MULTI-VIEW LEARNING

A Tabela 5 apresenta um resumo dos artigos, encontrados na literatura, que utilizam o conceito de MVL aplicados à SMC. A coluna ‘*Geração das views*’ apresenta as técnicas utilizadas para gerar as *views* em cada pesquisa. A coluna ‘*Classificadores*’ detalha os classificadores-base utilizados. A última coluna, ‘*Combinação*’, aborda as técnicas de combinação de classificadores utilizadas nos trabalhos.

Tabela 5 – Resumo das informações (método de geração de views, classificadores-base e métodos de combinação) relacionadas aos artigos sobre Multi-View Learning e Sistemas de Múltiplos Classificadores

Artigo	Ano	Geração das views	Classificadores	Combinação
(WANG; CHEN; SUN, 2007)	2007	<i>Kernel</i> linear, <i>Kernel</i> polinomial, <i>Kernel</i> RBF	SVM	Voto majoritário
(LE; KIM, 2016)	2016	Decomposição aleatória, Algoritmo Genético, Seleção de atributos, PCA	SVM	Voto ponderado
(KUMAR; MINZ, 2016)	2016	Partição do conjunto de características (heurística iterativa);	SVM, NB, k-NN	Voto ponderado
(LE; KIM, 2017)	2017	t-SNE, PCA	SVM	<i>Co-training</i>
(WANG et al., 2019)	2019	<i>Bi-Bagging</i> , <i>Fish Vector</i>	SVM, LR, Hokashyap;	Voto majoritário
(WANG et al., 2021)	2021	<i>Random Projection</i>	GB, LR, QDA, k-NN, AD, RF, XGBoost	Voto ponderado

Fonte: Elaborada pelo autor (2024)

3.4 DISCUSSÃO SOBRE OS TRABALHOS APRESENTADOS

As técnicas SDE apresentadas no início do capítulo (KNOP, KNORA-U, META-DES e DES-P), são interessantes de serem analisadas pela avaliação realizada na última revisão da literatura da área (CRUZ; SABOURIN; CAVALCANTI, 2018a), facilidade de reprodutibilidade dos resultados, com o código fonte disponível (CRUZ et al., 2020), além de serem comumente

utilizadas para avaliar novas abordagens SDE propostas (ZYBLEWSKI; SABOURIN; WOŹNIAK, 2021; ELMI; EFTEKHARI, 2021; FILHO; CANUTO; SANTIAGO, 2018; DO et al., 2020; WANG et al., 2021; CRUZ et al., 2019; CRUZ; SABOURIN; CAVALCANTI, 2018b; WOLOSZYNSKI et al., 2012).

Os trabalhos que utilizam *pools* homogêneos utilizam, majoritariamente, o classificador-base Árvores de decisão (AD), pela sua instabilidade e possibilidade de gerar mais diversidade ao *pool*. A maior parte dos trabalhos concentra esforços no âmbito da Seleção estática de classificadores. Na análise de trabalhos que utilizam *pools* formados por classificadores-base heterogêneos nota-se que não existe um padrão de tipos de classificadores utilizados. Algumas abordagens chegam a utilizar mais de 10 tipos de classificadores diferentes. As pesquisas também abordam as variadas fases dos SMC, como combinação e os dois tipos de seleção (estatística e dinâmica). Um ponto digno de nota é a falta de avaliação comparativa entre o desempenho das abordagens utilizando *pools* homogêneos e heterogêneos. Os trabalhos, em ambas as configurações de *pool*, quando abordam a seleção, fazem avaliação classificador-a-classificador, não avaliando se a composição dos classificadores é benéfica para o *ensemble*. Outro ponto de crítica é a falta de trabalhos que aproveitem a expertise adquirida pelas técnicas SDE em *ensembles* com o objetivo de aprimorar a classificação. Apenas o MLS utiliza diferentes técnicas SDE, contudo, sem a preocupação de avaliá-las conjuntamente.

Analisando os trabalhos apresentados na Tabela 5, dois pontos merecem destaque ao analisar esses trabalhos: **(i)** a ausência de trabalhos que utilizam combinações de diversas técnicas MVL para geração das *views*; **(ii)** a falta de trabalhos que aplicam seleção de classificadores. Sobre o primeiro ponto analisado, os trabalhos encontrados aplicam, geralmente, em seus experimentos apenas uma abordagem de geração de *views* (como modulação ou partição do conjunto de características). Esta tese supõe que combinar diversas técnicas pode aumentar a geração da diversidade no *pool*, já que não existe uma única representação ideal para qualquer base de dados. Sobre o segundo ponto, nenhum dos trabalhos aplicam a seleção de classificadores, seja estática ou dinâmica, focando apenas em combinação.

4 MÉTODO PROPOSTO

Este capítulo apresenta as abordagens propostas nesta tese. Inicialmente é apresentado o *Hardness-aware Oracle* (HaO) uma nova medida que visa avaliar os *ensembles* selecionados por múltiplas técnicas SDE. O HaO considera a análise de todos os classificadores presentes, em contraste com as atuais técnicas SDE que selecionam os classificadores que farão parte do *ensemble* sem avaliar o seu impacto na formação do *ensemble* como um todo. Além do HaO, também é apresentado o *Post-Selection Dynamic Ensemble Selection* (PS-DES), uma abordagem de SDE, que propõe selecionar os melhores *ensembles* gerados à partir de múltiplas técnicas SDE. Por último, é apresentando o *Post-Selection Dynamic Ensemble Selection with Multi-View Learning* (PS-DES-MVL), uma atualização do PS-DES que incorpora conceitos de *Multi-view learning*.

4.1 HARDNESS-AWARE ORACLE

No contexto da Seleção dinâmica de *ensemble* (SDE) as pesquisas estão focadas na proposição de novos métodos para selecionar os *ensembles*. Geralmente, esses métodos avaliam a competência de cada classificador individualmente, logo, não existe nenhuma forma de avaliação de *ensembles* gerados pelas técnicas SDE. Dado o exposto, e inspirado pelo Oráculo (SOUZA et al., 2017), é proposta uma forma de avaliar os *ensembles*, denominado *Hardness-aware Oracle* (HaO). Quando o Oráculo é aplicado como critério de avaliação, mesmo aplicado apenas para um classificador, torna-se superficial dado que que é focado na análise da presença de um único classificador competente. Em contraponto, o HaO considera a eficácia da seleção de classificadores competentes à partir do cálculo da proporção de classificadores competentes selecionados pelo *ensemble* em relação ao número total de classificadores competentes no *pool*.

O *Hardness-aware Oracle* é definido como a razão entre a quantidade de classificadores competentes, selecionados por uma técnica SDE pela quantidade total de classificadores selecionados pela mesma técnica. Dado uma amostra \mathbf{x} , dado um *pool* P com m classificadores-bases C , de forma que $P = \{C_1, C_2, \dots, C_m\}$; P'_x é o *ensemble* selecionado por uma técnica SDE para a amostra \mathbf{x} , de forma que $P'_x \subseteq P$; dado que O_x é o conjunto de classificadores competentes relacionados a classe de \mathbf{x} , definida por uma técnica SDE, para o *pool* P , de

forma que $O_x \subseteq P$, HaO_x é definido pela seguinte equação:

$$HaO_x(P_x, O_x) = \frac{|P'_x \cap O_x|}{|P'_x|} \quad (4.1)$$

Exemplificando: dado o *pool* P com 5 classificadores, de forma que $P = \{C_1, C_2, \dots, C_5\}$; dado uma amostra x_1 , o conjunto de classificadores competentes para x_1 (O_{x_1}) consiste em 4 classificadores: $|O_{x_1}| = 4$; dado que uma técnica de seleção obtém o *ensemble* P'_{x_1} com 4 classificadores, $|P'_{x_1}| = 4$, e sendo 3 desses classificadores competentes, $|P'_{x_1} \cap O_{x_1}| = 3$, HaO_{x_1} é calculado da seguinte forma:

$$HaO_{x_1}(P'_{x_1}, O_{x_1}) = \frac{|P'_{x_1} \cap O_{x_1}|}{|P'_{x_1}|} = \frac{3}{4} = 0,75 \quad (4.2)$$

Contudo, para a mesma amostra x_1 e para o mesmo *pool* P , se outra técnica de seleção de classificadores obtém o *ensemble* P''_{x_1} , com 4 classificadores competentes, $|P''_{x_1} \cap O_{x_1}| = 4$, HaO'_{x_1} é calculado da seguinte forma:

$$HaO'_{x_1}(P''_{x_1}, O_{x_1}) = \frac{|P''_{x_1} \cap O_{x_1}|}{|P''_{x_1}|} = \frac{4}{4} = 1 \quad (4.3)$$

Mesmo que ambas técnicas selecionem alguns classificadores competentes, para a amostra apresentada x_1 , como $HaO'_{x_1}(P'_{x_1}, O_{x_1}) < HaO'_{x_1}(P''_{x_1}, O_{x_1})$, a segunda técnica de seleção, que seleciona P''_{x_1} , é mais eficiente ao encontrar todos os classificadores competentes.

4.2 Post-Selection Dynamic Ensemble Selection

As diferentes abordagens para seleção de *ensembles* (e.g., *meta-aprendizado*, Oráculo, taxa de acerto) das técnicas SDE, geralmente, implicam na seleção de diferentes *ensembles*. Até o momento da escrita deste trabalho, não foram encontradas na literatura abordagens que possibilitem a utilização desses *ensembles*, em conjunto, objetivando melhorar o desempenho individual das abordagens SDE. O *Post-Selection Dynamic Ensemble Selection* (PS-DES) é uma abordagem que visa solucionar essa lacuna avaliando os *ensembles* gerados por múltiplas técnicas SDE à partir de uma novidade para a área: a fase de pós-seleção.

O *Post-Selection Dynamic Ensemble Selection* é composto de três fase: **(1)** geração do *pool* e preparação das técnicas SDE, que objetiva a geração de um *pool*, preferencialmente diverso, e a preparação do conjunto de técnicas SDE; **(2)** pós-seleção das técnicas SDE, que objetiva encontrar um *ensemble* ótimo para uma amostra apresentada; e **(3)** combinação, que objetiva combinar os classificadores no *ensemble* para classificar a amostra apresentada. As três fases do PS-DES são detalhadas nas seções a seguir.

4.2.1 PS-DES - Fase 1: Geração do *pool* e preparação das técnicas SDE

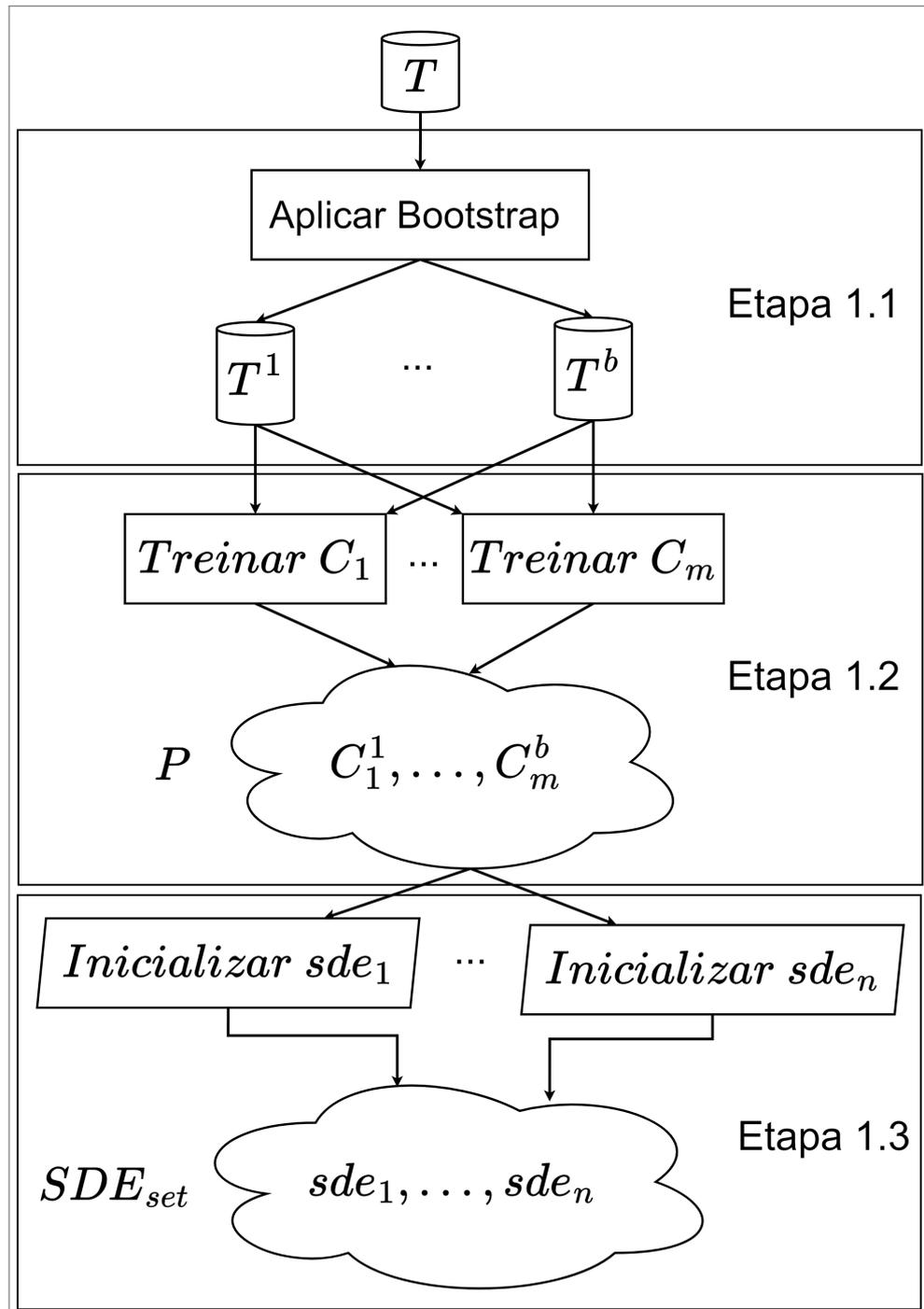
A primeira fase do PS-DES objetiva gerar um *pool* de classificadores e preparar um conjunto das técnicas SDE definido pelo usuário. A Figura 5 apresenta as três etapas do PS-DES. A Fase 1 é composta por 3 etapas. Inicialmente, na Etapa 1.1 é utilizado o *Bootstrap*. Em seguida, Etapa 1.2, todos os classificadores-base são treinados com cada uma das bases geradas na Etapa 1.1. Finalizando, Etapa 1.3, as técnicas SDE, definidas pelo usuário, são inicializadas para a próxima fase. As três etapas são detalhadas a seguir.

Etapa 1.1: Bootstrap. A aplicação do *Bootstrap*, similar ao Bagging, objetiva gerar diversidade entre os classificadores do *pool*. Para uma base de dados T , o *Bootstrap* é aplicado, com a quantidade de *bootstraps* (b), gerando assim b novas bases de dados. Exemplificando, aplicando um *Bootstrap* com $b = 2$, para a base de dados T , são gerados duas novas bases: T^1 e T^2 .

Etapa 1.2: Geração do *pool*. Nesta etapa, cada classificador C_i , de forma que $i = 1, 2, \dots, m$, onde m é a quantidade de classificadores-base, é treinado utilizando as bases de treino (T^1, \dots, T^b) gerados na etapa anterior. A saída desta etapa é o *pool* $P = \{C_1^1, C_1^2, \dots, C_m^b\}$, com $b \times m$ classificadores-base.

Etapa 1.3: Inicialização das técnicas SDE. O objetivo dessa etapa é a inicialização do conjunto de técnicas SDE, denominado SDE_{set} , de forma que $SDE_{set} = \{sde_1, sde_2, \dots, sde_n\}$, onde n é a quantidade de técnicas SDE definidas pelo usuário. Qualquer técnica SDE pode ser utilizada, contudo é interessante que elas utilizem diferentes critérios de seleção, para aumentar a probabilidade de gerar diferentes *ensembles* (e.g., META-DES, KNORA-U, DES-P). Cada técnica em SDE_{set} é inicializada utilizando o *pool* P .

Figura 4 – Fase 1 do PS-DES. O objetivo é gerar um *pool* de classificadores, preferencialmente diversos, e inicializar as técnicas SDE que serão utilizadas na próxima fase.



Fonte: Elaborada pelo autor (2024)

4.2.2 PS-DES - Fase 2: Seleção e Pós-seleção dinâmica

O objetivo da Fase 2 é obter, de forma dinâmica, um *ensemble* ótimo dado o conjunto de técnicas SDE SDE_{set} , definido na etapa anterior. Para uma amostra x_q , o conjunto de validação $DSEL$, o pool P e o conjunto de técnicas SDE, SDE_{set} , as quatro etapas da Fase

2 são especificadas da seguinte forma:

Etapa 2.1: Definição da Região de competência ($\theta_{\mathbf{x}_q}$). Dado a amostra \mathbf{x}_q e o conjunto de validação $DSEL$, a Região de competência é definida utilizando k-NN ou métodos de agrupamento. A RdC é utilizada pelas técnicas SDE para auxiliar na seleção dos *ensembles*. Dado que a definição da RdC é um dos gargalos computacionais nas técnicas SDE, objetivando mitigar esse custo é proposto o compartilhamento de uma única RdC para todas as técnicas SDE.

Etapa 2.2: Seleção de *ensembles* pelas técnicas SDE. O objetivo dessa etapa é gerar n *ensemble* P^n , escolhidos pela fase de seleção dinâmica de cada técnica SDE pertencente ao conjunto SDE_{set} . Esses *ensembles* são definidos à partir da RdC e da competência $\delta_{C_i, \mathbf{x}_q}$ de acordo com cada técnica SDE.

Etapa 2.3: Cálculo do Potencial do *ensemble*. O objetivo dessa etapa é avaliar cada um dos *ensembles* gerados pelas técnicas em SDE_{set} , gerados na etapa anterior, e identificar qual *ensemble* P^n é o mais apto para uma amostra \mathbf{x}_q . Para tal, é aplicado o conceito de Potencial, denominado POT_{P^n, \mathbf{x}_q} , uma forma de avaliar a qualidade do *ensemble*, à partir de diversas métricas, como acurácia, F-score e o HaO. Como a classe da amostra \mathbf{x}_q é desconhecida, assume-se que a classe é obtida pelo voto majoritário do *ensemble* P^n . Exemplificando: em um problema binário de classificação, e, dado um *ensemble* com 7 classificadores-base, definido por $P' = \{C_1, \dots, C_7\}$, selecionado por uma técnica SDE, e sendo $y_{P', \mathbf{x}_q} = \{0, 1, 0, 0, 1, 1, 1\}$ as classes definidas pelos classificadores-base em P' , para um padrão \mathbf{x}_q , o voto majoritário indicaria a classe 1 como resposta.

Etapa 2.4: Pós-seleção de *ensemble*. Uma das novidades propostas nesta tese. Essa etapa é responsável por escolher o *ensemble* P^{sel} , e, conseqüentemente, a técnica SDE, que obteve o maior valor de potencial, calculado na Etapa 2.3. Esta é a última etapa da Fase 2, o *ensemble* P^{sel} passa para a Fase 3, que visa integrar os classificadores. O Algoritmo 1 e a Figura 4 resumem o funcionamento da Fase 2.

O Algoritmo 1 recebe como parâmetros a amostra de teste (\mathbf{x}_q), a base de validação ($DSEL$) e o conjunto de técnicas SDE (SDE_{set}). Na linha é iniciada, com valor zero, a variável que representa o maior valor do Potencial encontrado (POT_{max}). Na linha é inicializado a variável que representa o melhor *ensemble* encontrado (P^{sel}). Na linha 4 é calculado a RdC para \mathbf{x}_q dado $DSEL$. Na linha 5 inicia-se uma iteração por todas as técnicas SDE presentes em SDE_{set} . Na linha 6, em P' é atribuído o *ensemble* selecionado por uma técnica *sde* para o padrão \mathbf{x}_q . Na linha 7 é calculado o Potencial de P' para o padrão \mathbf{x}_q . Na linha 8 é feito a

Algorithm 1 PS-DES seleção

```

1: procedure PS-DES_SELEÇÃO( $\mathbf{x}_q, DSEL, SDE_{set}$ )
2:    $POT_{max} \leftarrow 0$ 
3:    $P^{sel} \leftarrow \emptyset$ 
4:    $\theta_{\mathbf{x}_q} \leftarrow \text{calcular\_RdC}(\mathbf{x}_q, DSEL)$ 
5:   for  $des$  in  $SDE_{set}$  do
6:      $P' \leftarrow \text{get\_ensemble}(des, \theta_{\mathbf{x}_q})$ 
7:      $HaO_{des} \leftarrow \text{calcular\_POT}(P', \mathbf{x}_q)$ 
8:     if  $POT_{des} \geq POT_{max}$  then
9:        $POT_{max} \leftarrow POT_{des}$ 
10:       $P^{sel} \leftarrow P'$ 
11:    end if
12:  end for
13:  return  $P^{sel}$ 
14: end procedure

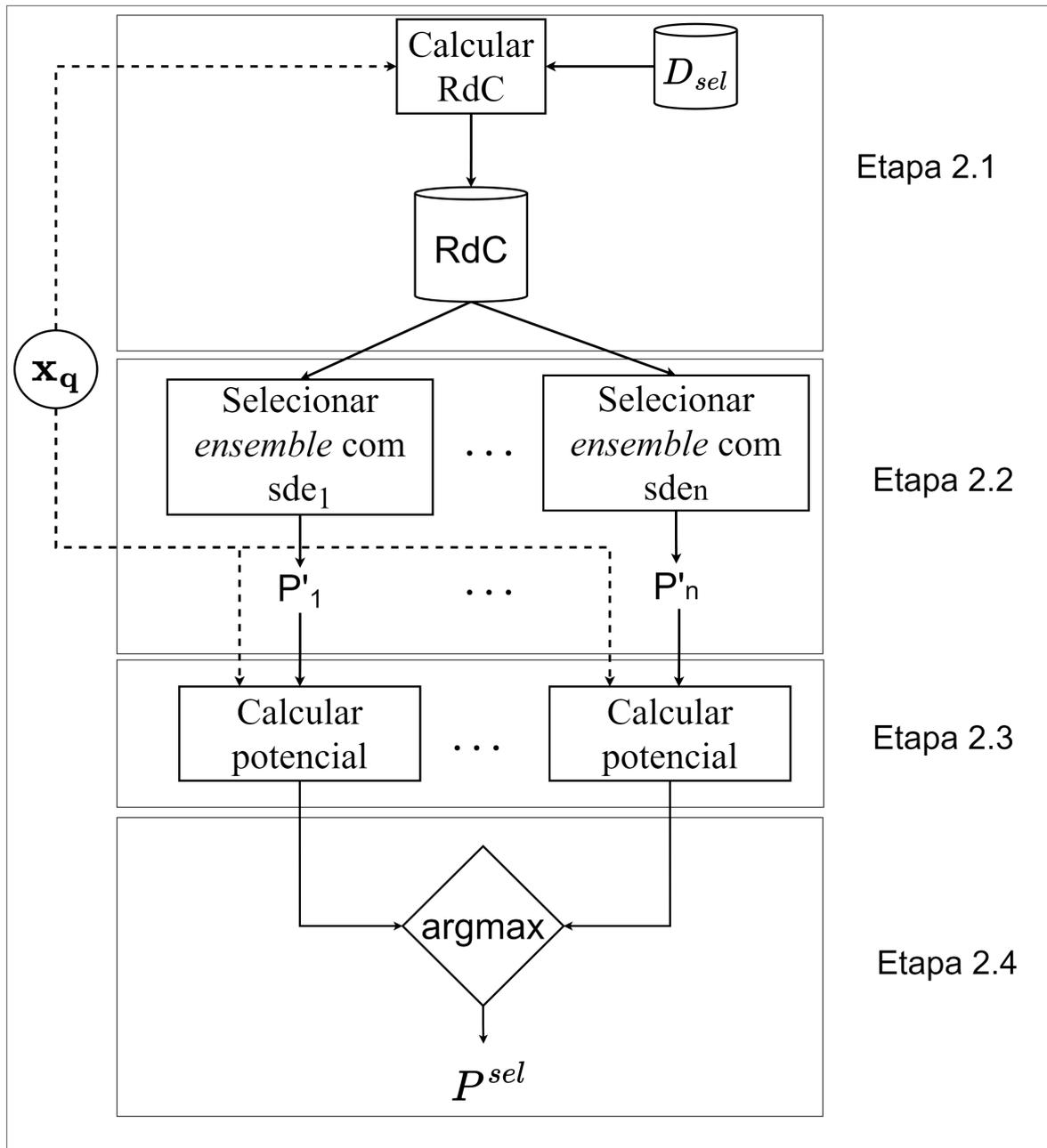
```

avaliação se o POT_{des} encontrado nessa iteração é maior ou igual do que o POT_{max} , e, em caso afirmativo, o POT_{max} e o P^{sel} são substituídos nas linhas 9 e 10. O algoritmo finaliza na linha 13 retornando o P^{sel} .

4.2.3 PS-DES - Fase 3: Combinação

A última fase é responsável por combinar os classificadores, caso mais de um classificador tenha sido selecionado, presentes no *ensemble* P^{sel} . Técnicas de combinação como voto majoritário ou a regra da soma podem ser utilizadas para este fim. A resposta é a classe da amostra \mathbf{x}_q .

Figura 5 – Fase 2 do PS-DES, seleção e pós-seleção dos *ensembles*. O objetivo dessa fase é selecionar o *ensemble* P^{sel} com maior valor de potencial.



Fonte: Elaborada pelo autor (2024)

4.3 Post-Selection Dynamic Ensemble Selection with Multi-View Learning

Partindo do pressuposto que a incorporação do MVL pode gerar mais diversidade ao sistema, é proposto adicionar uma nova camada ao PS-DES que aplica a geração de diversas *views*, para posteriormente selecionar qual *view* melhor se adapta para uma amostra apresentada. Essa versão do PS-DES é denominada *Post-Selection Dynamic Ensemble Selection with Multi-View Learning* (PS-DES-MVL).

O PS-DES-MVL tem quatro fases: (1) geração das *views* e *pool*, (2) seleção de *view*, (3) seleção e pós-seleção das técnicas SDE, (4) combinação. As fases serão detalhadas a seguir.

4.3.1 Fase 1: Geração das *views* e *pools*

Essa fase é responsável por gerar transformações na base de dado original, aplicando técnicas de geração de *views*, e posteriormente formar *pools* relativos a cada *view*. A Fase 1 é composta por quatro etapas:

Etapa 1.1: Geração das *views*. Nesta etapa, a base de treinamento (T) sofre transformações de dados (*i.e.*, são aplicadas técnicas de MVL), a partir de uma função (R), objetivando gerar novas bases diversas (*i.e.*, *views*), denominadas vT , sendo v a quantidade de funções escolhidas. As funções de transformação devem, preferencialmente, possuir diferentes propriedades (*e.g.*, espacial, *kernel* ou dissimilaridades), aumentando a probabilidade de gerar *views* mais diversas.

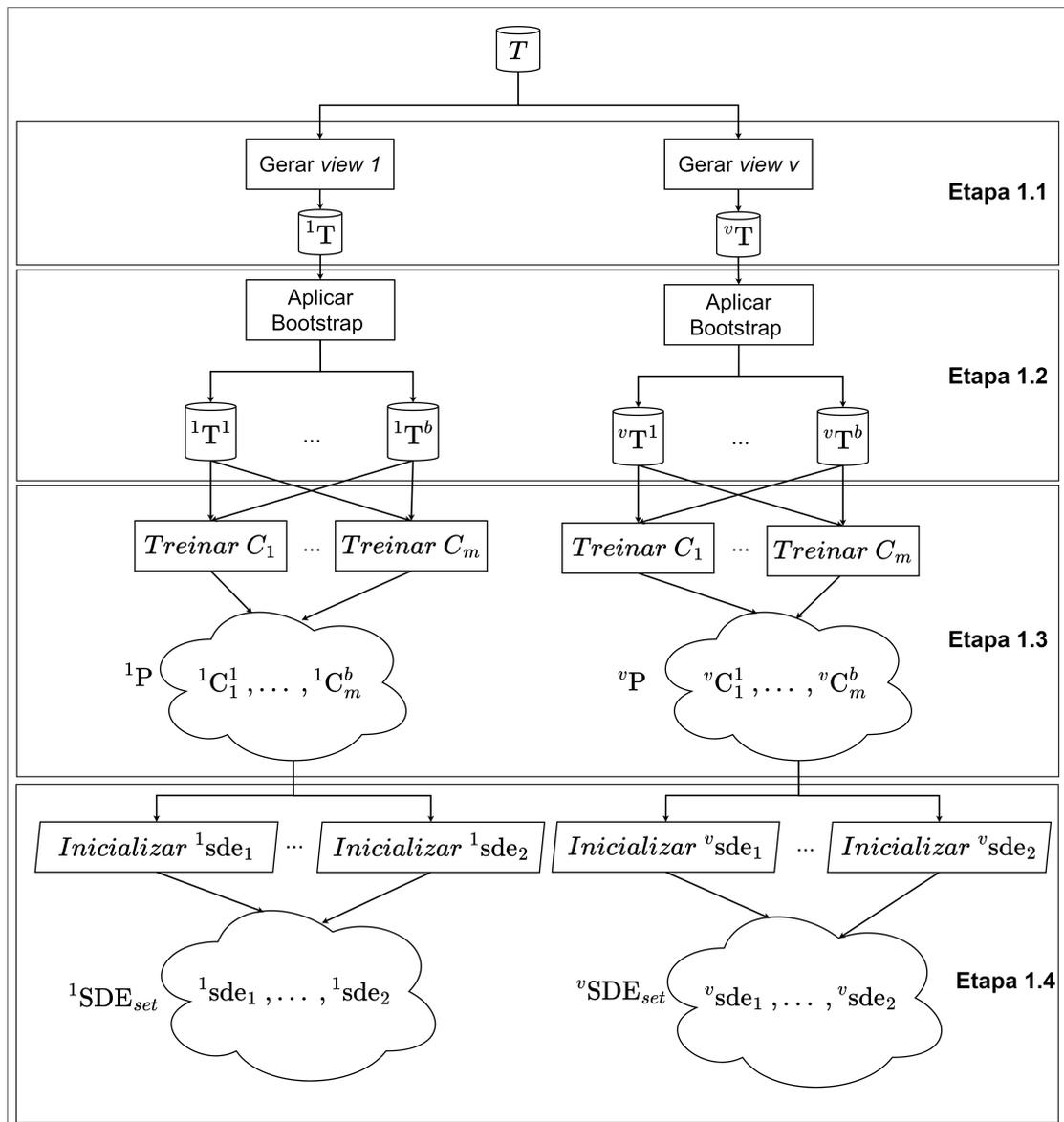
Etapa 1.2: *Bootstrap*. Essa etapa é semelhante a Etapa 1.1 do PS-DES. O objetivo é aplicar o *Bootstrap*, com b *bootstraps*, em cada uma das v bases geradas na etapa anterior, gerando assim $v \times b$ novas bases T . É importante frisar que as novas bases são geradas para cada uma das *views*. Exemplificando: utilizando duas funções de transformações, na Etapa 1.1 são geradas duas novas bases: 1T e 2T , em seguida, aplicando *Bootstrap* com $b = 2$, são gerados 4 novas bases, divididas em dois grupos de *views*: ${}^1T^1$ e ${}^1T^2$ para a *view* 1T , e as bases ${}^2T^1$ e ${}^2T^2$ relacionadas a 2T .

Etapa 1.3: Geração do *Pool*. Esta etapa é semelhante a Etapa 1.1 do PS-DES. Cada classificador C_i , de forma que $i = 1, 2, \dots, m$, onde m é a quantidade de classificadores-base, é treinado utilizando as bases de treino (${}^1T^1, \dots, {}^vT^b$) gerados na etapa anterior. Qualquer tipo de classificador pode ser utilizado como classificador-base. A saída desta etapa são v *pools* vP , cada um com $b \times m$ classificadores-base.

Etapa 1.4: Inicialização das técnicas SDE. Assim como na Etapa 1.3 do PS-DES, o objetivo dessa etapa é a inicialização de v conjuntos de técnicas SDE, denominados ${}^vSDE_{set}$, um conjunto para cada uma *views*, de forma que ${}^vSDE_{set} = \{{}^1sde_1, \dots, {}^v sde_n\}$, onde n é a quantidade de técnicas SDE definidas pelo usuário. Qualquer técnica SDE pode ser utilizada, contudo é interessante que elas utilizem diferentes critérios de seleção, para aumentar a probabilidade de gerar diferentes *ensembles* (*e.g.*, META-DES, KNORA-U, DES-P). Cada técnica em SDE_{set} é inicializada utilizando o *pool* vP .

A Figura 6 mostra a Fase 1 do PS-DES-MVL.

Figura 6 – Fase 1 do PS-DES-MVL



Fonte: Elaborada pelo autor (2024)

4.3.2 Fase 2: Seleção de *view*

Essa fase objetiva encontrar a melhor *view*, e, conseqüentemente, o melhor conjunto ${}^vSDE_{set}$ relacionado à *view*, gerado na fase anterior. Para definir qual a melhor *view*, utiliza-se o conceito de *k-Disagreeing Neighbors* (kDN) (SMITH; MARTINEZ; GIRAUD-CARRIER, 2014). Aplicado a seleção das *views*, o objetivo é escolher a *view* que obtém o menor valor de kDN de forma dinâmica, ou seja, para cada amostra apresentada. O motivo para essa escolha

baseia-se na seguinte hipótese: dado que cada *view* pode gerar diferentes RdC, ao escolher a RdC com menor probabilidade dos classificadores cometerem erros, aumenta a chance de uma classificação correta. Essa fase é composta de três etapas:

Etapa 2.1: definição da RdC para cada *view*. Como diferentes *views* podem gerar diferentes RdC (θ_{x_q}), é necessário gerar v distintas RdC, uma para cada representação de dados. A saída dessa fase é um conjunto de RdC, denominado RdC_{set} , de forma que $RdC_{set} = \{\theta_{x_q}^1, \dots, \theta_{x_q}^v\}$.

Etapa 2.2: cálculo do kDN. O objetivo é calcular o kDN, para a amostra x_q apresentada, dado uma θ_v e a base de treinamento relacionado à *view* 1T , utilizando a Equação 2.6. Como a classe de x_q não é conhecida, o kDN para cada *view* é calculado como a média dos valores de kDN para cada elemento de θ_v .

Etapa 2.3: seleção da *view*. O objetivo é selecionar a *view* que obteve o menor valor de kDN. A *view* selecionada será utilizada para a próxima fase.

O Algoritmo 2 exemplifica como funciona a Fase 2 do PS-DES-MVL.

Algorithm 2 Seleção de *view*

```

1: procedure SELEÇÃO_VIEW( $T_{set}, DSEL_{set}, x_{set}$ )
2:   for  $i$  in  $\{1, \dots, v\}$  do
3:      $\theta[i] \leftarrow \text{calcular\_RdC}(x_{set}[i], DSEL_{set}[i], 3)$ 
4:      $H[i] \leftarrow \text{kDN}(\theta[i], T_{set}[i])$ 
5:   end for
6:    $view_{index} \leftarrow \text{argmin}(H)$ 
7:   return  $view_{index}$ 
8: end procedure

```

O Algoritmo 2 inicia na linha 1, com os seguintes parâmetros: uma lista com bases de treinamentos, onde cada posição é relacionada a um das *views* (T_{set}), uma lista com as bases de validação, onde cada posição é relacionada a uma das *views* ($DSEL_{set}$), e uma lista com a amostra de teste, onde cada posição é a amostra relacionada a uma das *views* transformadas (x_{set}). Na linha 2 inicia-se um laço, onde para cada índice de *view* (i), é calculada a RdC, dado a amostra ($x_{set}[i]$) e a base de validação ($DSEL_{set}[i]$), e o resultado é armazenado em $\theta[i]$. Na linha 3 é calculado o kDN para $\theta[i]$ e $T_{set}[i]$. Na linha 6, é armazenado o índice da *view* que obteve o menor valor em H , sendo essa a *view* escolhida para a amostra apresentada. O resultado é retornado na linha 7. Caso várias *views* obtenham o mesmo valor de kDN, será retornada a primeira *view* que obteve o menor valor.

4.3.3 Fases 3: seleção e pós-seleção, e Fase 4: combinação

A Fase 3, seleção e pós-seleção das técnicas SDE, e a Fase 4, combinação, são, respectivamente, idênticas às Fases 2 e 3 do PS-DES.

5 ARRANJO EXPERIMENTAL E RESULTADOS

Este capítulo descreve o arranjo experimental utilizado para analisar o desempenho das abordagens propostas nesta tese, PS-DES e PS-DES-MVL, assim como os experimentos. São apresentadas as configurações das bases de dados, os modelos de aprendizagem de máquina e Seleção dinâmica de *ensemble* (SDE), assim como os seus parâmetros, as métricas para avaliação dos experimentos e as configurações das análises estatísticas. Os experimentos foram implementados em Python 3.8.5, usando a biblioteca de aprendizado de máquina Scikit-learn 1.0.1 (PEDREGOSA et al., 2011) e a biblioteca DESlib 0.3 (CRUZ et al., 2020).

5.1 ANÁLISES PROPOSTAS

Com o objetivo de avaliar o desempenho e o comportamento das abordagens propostas, foram elaborados experimentos, variando a formação do *pool* de classificadores, comparando a abordagem proposta com técnicas SDE individuais, além da aplicação do MVL. O objetivo dos experimentos é realizar as seguintes análises:

- **Análise 1: Avaliação do PS-DES utilizando *pools* homogêneos.** O objetivo é avaliar o desempenho do PS-DES, com o HaO como forma de calcular o potencial, utilizando *pools* homogêneos, quando comparado a técnicas SDE individuais utilizando os mesmos *pools*, além de uma abordagem de pós-seleção que escolhe aleatoriamente uma das técnicas SDE definidas, denominada PS-Aleatório, que utiliza as mesmas configurações do PS-DES.
- **Análise 2: Avaliação do PS-DES utilizando *pools* heterogêneos.** O objetivo é avaliar o desempenho do PS-DES, com o HaO como forma de calcular o potencial, utilizando *pools* formados por diferentes modelos de aprendizagem de máquina, quando comparado a técnicas SDE individuais e ao PS-Aleatório, utilizando esses mesmos *pools*.
- **Análise 3: Avaliação da diversidade de *pools* homogêneos e heterogêneos.** O objetivo é avaliar a diversidade entre configurações de *pools* homogêneos e heterogêneos.
- **Análise 4: Avaliação do PS-DES utilizando *pool* heterogêneo vs homogêneo.** O objetivo é avaliar o desempenho do PS-DES, com o HaO como forma de calcular o

potencial, configurado com um *pool* heterogêneo quando comparado ao próprio PS-DES utilizando o PS-DES configurado com três *pools* homogêneos.

- **Análise 5: Avaliação do PS-DES utilizando diferentes formas para calcular o potencial.** O objetivo é avaliar o desempenho do PS-DES, configurado com um *pool* heterogêneo, utilizando três métricas para calcular o potencial comparadas ao PS-DES utilizando o HaO para avaliar o potencial.
- **Análise 6: Comportamento de seleção do PS-DES.** O objetivo é avaliar as técnicas SDE escolhidas pelo PS-DES, com o HaO como forma de calcular o potencial, objetivando entender o funcionamento da seleção na abordagem proposta.
- **Análise 7: Avaliação do PS-DES-MVL com *pools* heterogêneos.** O objetivo é avaliar, de forma comparativa, o comportamento do PS-DES-MVL, com o HaO como forma de calcular o potencial, com as técnicas SDE individuais quando alimentadas com diferentes técnicas de geração de *views*, utilizadas de forma individual e combinadas.

5.2 BASES DE DADOS

Primeiro, o conjunto de dados T foi dividido em três partes: 50% para treinamento, 25% para validação e 25% para teste. Toda a divisão foi realizada de forma estratificada, o que significa que as proporções das classes entre os conjuntos de treinamento e teste foram mantidas. Em seguida, cada uma das bases, separadamente, tiveram suas amostras normalizadas, usando o *Standard Scaler*, também conhecido como normalização-Z (AMORIM; CAVALCANTI; CRUZ, 2023). Isto significa que as variáveis foram transformadas para que tenham média zero e desvio padrão igual a um. Para a Análise 6, também foram aplicadas as funções de geração de *views* nas bases de validação e teste.

Os experimentos foram realizados com 30 conjuntos de dados do UCI Machine Learning Repository (DUA; GRAFF, 2020). As bases apresentam diferentes características, como número de amostras, dimensões, classes e a Taxa de desbalanceamento (TD), calculado pela razão entre as cardinalidades da classe com mais amostras e a cardinalidade da classe com menos amostras. As características de cada base de dados são apresentadas na Tabela 6. Foram executadas trinta repetições para cada conjunto de dados, alterando a distribuição dos conjuntos (*holdout*) para obter os valores médios das métricas avaliadas.

Tabela 6 – Descrição das bases de dados: quantidade de amostras, dimensões, classes, e taxa de desbalanceamento (TD).

Base	Amostras	Dimensões	Classes	TD
appendicitis	106	7	2	2,52
banana	5300	2	2	1,00
blood	748	4	2	3,20
bupa	345	6	2	1,38
cleveland	297	13	5	12,31
cmc	1473	9	3	1,30
column_3C	310	6	3	2,50
dermatology	358	34	6	5,55
diabetes	768	8	2	1,86
ecoli-0_vs_1	220	7	2	1,86
german	1000	20	2	3,33
glass1	214	9	2	1,82
glass6	214	9	2	6,38
haberman	306	3	2	2,78
hayes-Roth	160	4	3	3,40
heart	270	13	2	1,25
ilpd	583	10	2	2,50
led7digit	500	7	10	1,54
musk	476	166	2	1,29
new-thyroid1	215	5	2	5,14
pima	768	8	2	1,90
sonar	208	60	2	1,14
transfusion	400	4	2	2,57
vehicle	846	18	4	1,10
vehicle2	846	18	2	2,88
vowel	990	13	11	1,00
vowel0	988	13	2	9,95
wdbc	683	9	2	1,85
wine	178	13	3	1,00
yeast-2_vs_4	514	8	2	9,08

Fonte: Elaborada pelo autor (2024)

5.3 PARÂMETROS DOS ALGORITMOS UTILIZADOS

Foram utilizados três tipos de classificadores-base para execução dos experimentos: NB, LR e Perceptron. Esses classificadores foram escolhidos por terem diferentes bases matemáticas além de baixo custo computacional (NGUYEN et al., 2020; COSTA et al., 2018; CRUZ; SABOURIN; CAVALCANTI, 2018a). Como a otimização dos hiperparâmetros não é um dos objetivos do trabalho, além de tentar facilitar a reprodução dos experimentos, classificadores-base utilizam os hiperparâmetros padrões da biblioteca sklearn (PEDREGOSA et al., 2011). Para as configurações de *pools* heterogêneos, nas quais são combinados os três classificadores-base, foram utilizados 100 *bootstrap* para cada classificador-base, de forma similar a outros trabalhos (CRUZ; SABOURIN; CAVALCANTI, 2018a; KO; SABOURIN; JR, 2008; CRUZ et al., 2015), totalizando assim um *pool* com 300 classificadores. Para manter os *pools* com o mesmo tamanho, para as configurações de classificadores homogêneos, são utilizados 300 *bootstraps*.

Foram selecionadas quatro técnicas SDE: KNOP, KNORA-U, DES-P e META-DES. Os métodos selecionados foram escolhidos por apresentarem conceitos de seleção distintos, reprodutibilidade e por serem comumente utilizado para comparar novas abordagens SDE. Além disso, apresentam melhor desempenho, de acordo com o último estudo comparativo da área (CRUZ; SABOURIN; CAVALCANTI, 2018a). As técnicas foram configuradas usando as configurações de hiperparâmetros padrões da biblioteca DESlib (CRUZ et al., 2020). Para uma comparação justa, as técnicas SDE usam o mesmo tamanho do *pool* e os mesmos parâmetros para os classificadores-base.

Para o PS-DES é necessário definir os parâmetros e modelos para as três fases. Na Fase 1, geração do *pool* e preparação das técnicas SDE, foram utilizados os três classificadores-base descritos anteriormente (NB, LR e Perceptron). As técnicas individuais SDE que foram utilizadas nesta fase também são as descritas anteriormente: KNOP, KNORA-U, DES-P e META-DES. Os mesmos parâmetros para as avaliações individuais das classificadores-base e técnicas individuais SDE são utilizados quando aplicados ao PS-DES. Na Fase 2, seleção, o único parâmetro necessário de configuração é a definição da RdC, calculada utilizando o k-NN, com número de vizinhos igual a 7, seguindo um valor comum na literatura (CRUZ; SABOURIN; CAVALCANTI, 2018a). Na última fase, combinação, é utilizado o voto majoritário como abordagem para integrar o *ensemble* selecionado.

Para o PS-DES-MVL (Análise 6): na Fase 1, geração das *views* e *pools*, foram utilizados as cinco técnicas de geração de *view* (t-SNE, PCA Kernel Linear, Kernel Polinomial e Kernel

RBF). A configuração do experimento para os classificadores-base, *bootstraps* e as técnicas individuais SDE são as mesmas apresentadas para o PS-DES, logo com a mesma quantidade de classificadores no *pool*, para cada *view*. Para as técnicas de geração de *views* são utilizados os hiperparâmetros padrões do scikit-learn (PEDREGOSA et al., 2011). Na Fase 2 e Fase 3, seleção de *view*, o único parâmetro necessário de configuração é a definição da RdC, calculada utilizando o k-NN, com número de vizinhos igual a 7, seguindo um valor comum na literatura (CRUZ; SABOURIN; CAVALCANTI, 2018a). Na Fase 4, combinação, é utilizado o voto majoritário como abordagem para integrar o *ensemble* selecionado.

5.4 MÉTRICAS PARA AVALIAÇÃO DE DESEMPENHO

Foram selecionadas três métricas para comparar o desempenho das técnicas utilizadas nos análises propostas: acurácia, F-score e *Matthews Correlation Coefficient* (MCC). A acurácia é uma das métricas mais comumente utilizadas na avaliação de desempenho de técnicas de aprendizagem de máquina, contudo, não é eficaz para avaliar bases de dados desbalanceadas. Como algumas das bases de dados utilizadas nos experimentos são desbalanceadas foram utilizadas as métricas F-score e *Matthews Correlation Coefficient*, que são mais indicadas para esse tipo de situação. As métrica são definidas pelas seguintes equações:

$$\text{Acurácia} = \frac{VP + TF}{VP + TF + FP + FN} \quad (5.1)$$

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (5.2)$$

$$\text{Recall} = \frac{VP}{FN + VP} \quad (5.3)$$

$$\text{F-score} = 2 \times \frac{\text{Precisão} \times \text{Recall}}{\text{Precisão} + \text{Recall}} \quad (5.4)$$

$$\text{MCC} = \frac{(VP \times VN) - (FP \times FN)}{\sqrt{(VP + FP)(VP + FN)(VN + FP) + (VN + FN)}} \quad (5.5)$$

onde, *VP* é o verdadeiro positivo, *VN* é o verdadeiro negativo, *FP* é o falso positivo, e *FN* é o falso negativo.

5.5 ANÁLISE ESTATÍSTICA

Com o objetivo de gerar análises mais robustas e confiáveis, além da apresentação das médias das métricas de avaliação, foram aplicados testes estatísticos em todas as análises, seguindo o processo proposto em (BENAVOLI; CORANI; MANGILI, 2016):

1. **Aplicação do teste de hipótese de Friedman:** o objetivo desse primeiro teste é avaliar se existe diferença entre as amostras do experimento avaliado, considerando as múltiplas abordagens e as todas as bases de dados. Para isto é preciso avaliar se a hipótese nula é rejeitada, ou seja, o p-valor é menor do que o nível de confiança estipulado. Como todos os p-valores do teste de Friedman foram muito próximos a zero para todas as análises desta tese, esses valores não serão apresentados nos resultados.
2. **Aplicação do teste de hipóteses pareado de Wilcoxon:** caso a hipótese nula seja rejeitada no teste de Friedman, o próximo passo é avaliar, par-a-par todas as abordagens propostas nas análises utilizando uma abordagem como controle. Nesta situação, a hipótese nula discute se a abordagem de controle obteve resultados superiores quando comparada a uma das outras abordagens apresentadas nas análises. Para todos os testes estatísticos apresentados durante esta tese, é utilizado o nível de confiança (α) em 95%.

5.6 ANÁLISE 1: AVALIAÇÃO DO PS-DES UTILIZANDO *POOLS* HOMOGÊNEOS

Na primeira análise foi avaliado o desempenho da abordagem proposta nesta tese, PS-DES, utilizando três configurações de *pools* homogêneos, quando comparados a técnicas SDE e o PS-Aleatório. Os *pools* foram compostos por três classificadores-base: NB, com os resultados apresentados na Tabela 7, LR, com os resultados apresentados na Tabela 8 e Perceptron, com os resultados apresentados na Tabela 9. As tabelas apresentam os resultados da análise estatística do teste de Wilcoxon, comparando o PS-DES com as abordagens SDE, utilizando as 30 bases de dados apresentadas. Os resultados de média para cada configuração de *pool* e métrica avaliada, podem ser encontrados no Apêndice A.

Analisando a Tabela 7 o PS-DES obteve melhores resultados de ranking quando comparado a todas as abordagens SDE e ao PS-Aleatório. Ao analisar o p-valor, os resultados sugerem que o PS-DES é superior ao KNOP, KNORA-U, DES-P e ao PS-Aleatório nas três métricas

Tabela 7 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro abordagens SDE (KNOP, KNORA-U, META-DES, DES-P) e o PS-Aleatório. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores NB.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
Acurácia	ranking	4,12	4,75	2,70	3,10	3,95	2,38
	p-valor	0,000	0,000	0,149	0,012	0,000	-
F-score	ranking	4,12	4,75	2,70	3,10	3,95	2,38
	p-valor	0,000	0,000	0,149	0,012	0,000	-
MCC	ranking	4,18	4,42	3,00	2,93	3,90	2,57
	p-valor	0,002	0,000	0,154	0,038	0,000	-

Fonte: Elaborada pelo autor (2024)

avaliadas. Os resultados foram estatisticamente equivalentes ao META-DES nas três métricas avaliadas.

Tabela 8 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro abordagens SDE (KNOP, KNORA-U, META-DES, DES-P) e o PS-Aleatório. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores LR.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
Acurácia	ranking	4,32	4,33	2,53	3,35	3,80	2,67
	p-valor	0,002	0,001	0,573	0,068	0,002	-
F-score	ranking	4,48	4,87	2,32	3,05	3,90	2,38
	p-valor	0,000	0,000	0,444	0,003	0,000	-
MCC	ranking	4,45	4,50	2,48	3,22	3,80	2,55
	p-valor	0,000	0,000	0,444	0,008	0,000	-

Fonte: Elaborada pelo autor (2024)

Na análise da Tabela 8 e, PS-DES obteve o menor ranking entre quase todas as abordagens avaliadas, perdendo apenas para o META-DES. Na análise do p-valor, a abordagem proposta obteve melhores resultados em todas as métricas quando comparadas ao KNOP, KNORA-U, DES-P e ao PS-Aleatório. Os resultados foram estatisticamente equivalentes ao META-DES nas três métricas avaliadas.

Na última análise dos experimentos relacionados à configurações que utilizam *pool* homogêneo de Perceptrons, Tabela 9, o PS-DES novamente obteve melhores resultados de ranking, para todas as três métricas avaliadas. No tocante ao p-valor, comparado as técnicas KNOP, KNORA-U, DES-P e ao PS-Aleatório os resultados indicam que o PS-DES obteve resulta-

Tabela 9 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro abordagens SDE (KNOP, KNORA-U, META-DES, DES-P) e o PS-Aleatório. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores Perceptron.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
Acurácia	ranking	4,30	4,93	2,93	2,90	3,82	2,12
	p-valor	0,000	0,000	0,107	0,020	0,000	-
F-score	ranking	4,57	5,23	2,33	3,13	3,83	1,90
	p-valor	0,000	0,000	0,329	0,000	0,000	-
MCC	ranking	4,33	5,00	2,77	2,87	4,03	2,00
	p-valor	0,000	0,000	0,111	0,007	0,000	-

Fonte: Elaborada pelo autor (2024)

dos estatisticamente superiores. Assim como nas outras configurações de *pool*, os resultados indicam equivalência estatística entre o PS-DES e o META-DES.

Resumindo os resultados apresentados para esta análise: o PS-DES obteve melhor desempenho, nas métricas avaliadas, quando comparado as abordagens KNOP, KNORA-U, DES-P e ao PS-Aleatório, e resultados estatisticamente equivalentes ao META-DES, o que pode ser explicado pelo desempenho superior do META-DES, como apresentada na literatura (CRUZ; SABOURIN; CAVALCANTI, 2018a), portanto, o PS-DES fica limitado ao desempenho dessa abordagem, dado que o META-DES é selecionado na maioria das vezes.

5.7 ANÁLISE 2: AVALIAÇÃO DO PS-DES UTILIZANDO *POOL* HETEROGÊNEO

Este experimento trata da avaliação comparativa do PS-DES quando comparado a quatro técnicas SDE (KNOP, KNORA-U, META-DES e DES-P) e ao PS-Aleatório, utilizando um *pool* heterogêneo com três classificadores-base: NB, LR e Perceptron. Os resultados dos experimentos estão separados pela métrica utilizada: acurácia (Tabela 10), F-score (Tabela 11) e MCC (Tabela 12). As tabelas apresentam o valor médio para cada abordagem e base de dados. Ao final, também é apresentado o ranking médio e o resultado (p-valor) do teste pareado de Wilcoxon.

Os resultados das Tabelas 10, 11, e 12 mostram que o PS-DES obteve, para todas as métricas, o melhor ranking dentre as abordagens avaliadas. Avaliando o p-valor, é possível afirmar que o PS-DES é estatisticamente superior em termos de acurácia, F-score e MCC. Portanto, o esquema de seleção de pós-seleção proposto resulta em uma melhoria significativa

Tabela 10 – Avaliação da acurácia média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* heterogêneo de três classificadores básicos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,862	0,859	0,851	0,864	0,855	0,865
banana	0,690	0,691	0,752	0,748	0,721	0,749
blood	0,778	0,779	0,776	0,784	0,781	0,787
bupa	0,668	0,662	0,640	0,662	0,654	0,667
cleveland	0,621	0,619	0,618	0,619	0,623	0,614
cmc	0,512	0,509	0,488	0,513	0,506	0,513
column_3C	0,844	0,846	0,841	0,848	0,846	0,845
dermatology	0,972	0,972	0,970	0,972	0,970	0,971
diabetes	0,769	0,768	0,763	0,765	0,767	0,766
ecoli-0_vs_1	0,971	0,970	0,972	0,969	0,970	0,972
german	0,741	0,742	0,740	0,744	0,742	0,744
glass1	0,640	0,642	0,674	0,682	0,653	0,682
glass6	0,938	0,940	0,947	0,944	0,943	0,947
haberman	0,731	0,731	0,726	0,737	0,734	0,733
hayes	0,619	0,611	0,670	0,640	0,640	0,685
heart	0,838	0,836	0,833	0,841	0,836	0,839
ilpd	0,708	0,710	0,688	0,682	0,700	0,687
led7digit	0,726	0,725	0,708	0,723	0,721	0,721
musk	0,781	0,779	0,788	0,792	0,786	0,798
new-thyroid1	0,946	0,947	0,956	0,948	0,949	0,956
pima	0,768	0,770	0,761	0,766	0,766	0,765
sonar	0,771	0,771	0,796	0,784	0,786	0,799
transfusion	0,778	0,780	0,780	0,785	0,780	0,785
vehicle	0,752	0,750	0,761	0,752	0,750	0,762
vehicle2	0,947	0,949	0,956	0,951	0,951	0,955
vowel	0,965	0,965	0,981	0,967	0,969	0,978
vowel0	0,965	0,965	0,980	0,967	0,969	0,978
wdbc	0,970	0,970	0,969	0,970	0,970	0,970
wine	0,974	0,971	0,975	0,974	0,973	0,975
yeast-2_vs_4	0,942	0,942	0,942	0,946	0,940	0,945
ranking	4,03	4,10	3,88	2,90	3,73	2,35
p-value	0,002	0,001	0,000	0,007	0,000	-

Fonte: Elaborada pelo autor (2024)

Tabela 11 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* heterogêneo de três classificadores básicos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,718	0,708	0,717	0,749	0,715	0,758
banana	0,635	0,636	0,723	0,720	0,682	0,721
blood	0,531	0,531	0,568	0,604	0,565	0,610
bupa	0,619	0,610	0,618	0,634	0,616	0,641
cleveland	0,343	0,344	0,357	0,349	0,352	0,339
cmc	0,477	0,473	0,454	0,484	0,472	0,483
column_3C	0,796	0,799	0,792	0,803	0,799	0,798
dermatology	0,968	0,968	0,965	0,968	0,966	0,966
diabetes	0,725	0,725	0,719	0,723	0,724	0,726
ecoli-0_vs_1	0,967	0,966	0,969	0,965	0,966	0,968
german	0,623	0,622	0,639	0,638	0,630	0,645
glass1	0,415	0,422	0,585	0,577	0,512	0,557
glass6	0,847	0,850	0,880	0,868	0,863	0,876
haberman	0,507	0,506	0,525	0,542	0,519	0,554
hayes	0,634	0,627	0,688	0,658	0,654	0,701
heart	0,834	0,832	0,829	0,837	0,832	0,835
ilpd	0,502	0,503	0,586	0,553	0,549	0,570
led7digit	0,719	0,718	0,700	0,715	0,714	0,716
musk	0,772	0,770	0,781	0,785	0,778	0,791
new-thyroid1	0,885	0,887	0,909	0,890	0,893	0,908
pima	0,724	0,727	0,718	0,724	0,723	0,724
sonar	0,767	0,768	0,793	0,781	0,783	0,795
transfusion	0,529	0,533	0,572	0,606	0,562	0,613
vehicle	0,745	0,742	0,758	0,744	0,743	0,757
vehicle2	0,931	0,933	0,943	0,936	0,935	0,942
vowel	0,886	0,886	0,940	0,894	0,899	0,931
vowel0	0,885	0,885	0,939	0,893	0,898	0,932
wdbc	0,966	0,966	0,965	0,967	0,967	0,967
wine	0,974	0,971	0,975	0,974	0,973	0,975
yeast-2_vs_4	0,786	0,785	0,787	0,804	0,778	0,800
ranking	4,43	4,60	3,00	2,87	4,10	2,00
p-value	0,000	0,000	0,024	0,002	0,000	-

Fonte: Elaborada pelo autor (2024)

Tabela 12 – Avaliação do MCC médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* heterogêneo de três classificadores básicos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,517	0,503	0,489	0,563	0,500	0,571
banana	0,428	0,435	0,538	0,525	0,484	0,528
blood	0,196	0,199	0,220	0,275	0,229	0,284
bupa	0,309	0,293	0,250	0,295	0,272	0,306
cleveland	0,358	0,353	0,357	0,356	0,362	0,345
cmc	0,233	0,227	0,196	0,238	0,224	0,237
column_3C	0,753	0,755	0,748	0,759	0,755	0,754
dermatology	0,965	0,966	0,963	0,965	0,963	0,964
diabetes	0,468	0,466	0,454	0,460	0,464	0,465
ecoli-0_vs_1	0,937	0,936	0,940	0,934	0,935	0,939
german	0,314	0,314	0,319	0,328	0,320	0,330
glass1	0,008	0,019	0,226	0,242	0,134	0,225
glass6	0,723	0,730	0,770	0,755	0,749	0,768
haberman	0,130	0,131	0,144	0,193	0,157	0,187
hayes	0,427	0,416	0,505	0,467	0,461	0,529
heart	0,677	0,673	0,667	0,682	0,673	0,677
ilpd	0,116	0,131	0,188	0,144	0,155	0,173
led7digit	0,697	0,696	0,678	0,693	0,692	0,692
musk	0,551	0,547	0,566	0,575	0,562	0,586
new-thyroid1	0,797	0,800	0,834	0,806	0,807	0,832
pima	0,466	0,471	0,451	0,462	0,461	0,461
sonar	0,544	0,547	0,594	0,570	0,576	0,600
transfusion	0,192	0,204	0,232	0,277	0,227	0,286
vehicle	0,673	0,672	0,685	0,674	0,671	0,686
vehicle2	0,863	0,867	0,886	0,873	0,872	0,884
vowel	0,782	0,784	0,885	0,798	0,807	0,867
vowel0	0,780	0,780	0,883	0,796	0,806	0,870
wdbc	0,933	0,933	0,931	0,934	0,933	0,934
wine	0,962	0,957	0,963	0,961	0,960	0,963
yeast-2_vs_4	0,632	0,629	0,635	0,657	0,618	0,653
ranking	4,27	4,43	3,37	2,77	3,93	2,23
p-value	0,000	0,000	0,011	0,004	0,000	-

Fonte: Elaborada pelo autor (2024)

em relação aos métodos existentes. É possível que a diversidade dos *pools* heterogêneos tenha relação com a eficácia da proposta, por isso, na próxima análise é investigada a diversidade entre os *pools* homogêneos e heterogêneos.

5.8 ANÁLISE 3: AVALIAÇÃO DA DIVERSIDADE DE *POOLS* HOMOGÊNEOS E HETEROGÊNEOS

Os resultados apresentados nessa seção mostram a diversidade entre as configurações de *pool* com classificadores homogêneos e heterogêneos. Foram utilizadas duas métricas para calcular a diversidade: *Disagreement*, com os resultados apresentados na Tabela 13, e *Double-Fault*, com os resultados apresentados na Tabela 14.

Para as duas métricas avaliadas, nas análises estatísticas pareadas (Tabela 13 e 14) o *pool* formado por classificadores heterogêneos obteve o menor ranking e resultados estatisticamente superiores as configurações com *pools* homogêneos. Logo, pode-se concluir que, para as duas métricas avaliadas, a abordagem heterogênea é mais diversa.

Tabela 13 – Avaliação da métrica de diversidade Disagreement comparando um *pool* heterogêneo com três configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado para cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado o *pool* heterogêneo com a configuração de *pools* homogêneos apresentada na coluna.

Base	NB	LR	Perceptron	Heterogêneo
appendicitis	0,071	0,067	0,061	0,111
banana	0,061	0,019	0,024	0,040
blood	0,041	0,029	0,021	0,054
bupa	0,299	0,199	0,226	0,337
cleveland	0,191	0,090	0,095	0,156
cmc	0,129	0,155	0,196	0,301
column_3C	0,115	0,083	0,100	0,131
dermatology	0,052	0,030	0,038	0,103
diabetes	0,083	0,075	0,130	0,131
ecoli-0_vs_1	0,103	0,020	0,034	0,092
german	0,118	0,102	0,070	0,161
glass1	0,245	0,142	0,119	0,272
glass6	0,082	0,045	0,052	0,073
haberman	0,045	0,052	0,040	0,062
hayes	0,298	0,325	0,382	0,358
heart	0,092	0,107	0,142	0,129
ilpd	0,053	0,099	0,046	0,311
led7digit	0,249	0,130	0,165	0,349
musk	0,109	0,136	0,175	0,196
new-thyroid1	0,024	0,019	0,020	0,043
pima	0,082	0,074	0,130	0,130
sonar	0,158	0,196	0,220	0,237
transfusion	0,042	0,028	0,021	0,054
vehicle	0,188	0,121	0,163	0,267
vehicle2	0,129	0,027	0,049	0,120
vowel	0,039	0,020	0,029	0,050
vowel0	0,038	0,020	0,029	0,050
wdbc	0,019	0,014	0,025	0,031
wine	0,036	0,032	0,043	0,044
yeast-2_vs_4	0,132	0,019	0,025	0,083
ranking	2,33	3,53	2,87	1,27
p-value	0,000	0,000	0,000	-

Fonte: Elaborada pelo autor (2024)

Tabela 14 – Avaliação da métrica de diversidade Double-fault comparando um *pool* heterogêneo com três configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado para cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado o *pool* heterogêneo com a configuração de *pools* homogêneos apresentada na coluna.

Base	NB	LR	Perceptron	Heterogêneo
appendicitis	0,098	0,134	0,152	0,106
banana	0,102	0,114	0,112	0,106
blood	0,218	0,205	0,217	0,201
bupa	0,299	0,253	0,276	0,228
cleveland	0,330	0,347	0,362	0,331
cmc	0,470	0,424	0,450	0,377
column_3C	0,150	0,122	0,119	0,115
dermatology	0,170	0,023	0,024	0,041
diabetes	0,213	0,198	0,201	0,187
ecoli-0_vs_1	0,096	0,031	0,025	0,030
german	0,217	0,210	0,253	0,194
glass1	0,295	0,303	0,312	0,251
glass6	0,039	0,046	0,070	0,045
haberman	0,238	0,242	0,250	0,235
hayes	0,245	0,283	0,277	0,256
heart	0,142	0,143	0,140	0,133
ilpd	0,392	0,243	0,267	0,179
led7digit	0,663	0,235	0,243	0,296
musk	0,217	0,185	0,179	0,166
new-thyroid1	0,053	0,049	0,052	0,040
pima	0,213	0,198	0,202	0,187
sonar	0,240	0,157	0,156	0,162
transfusion	0,217	0,206	0,216	0,201
vehicle	0,440	0,197	0,180	0,217
vehicle2	0,134	0,047	0,040	0,048
vowel	0,046	0,030	0,036	0,027
vowel0	0,046	0,030	0,036	0,027
wdbc	0,037	0,029	0,028	0,025
wine	0,025	0,021	0,020	0,019
yeast-2_vs_4	0,083	0,046	0,053	0,050
ranking	3,30	2,43	2,70	1,57
p-value	0,000	0,003	0,002	-

Fonte: Elaborada pelo autor (2024)

5.9 ANÁLISE 4: AVALIAÇÃO DO PS-DES UTILIZANDO *POOL* HETEROGÊNEO VS HOMOGÊNEO.

Como o PS-DES obteve os melhores resultados nas análises com *pools* homogêneos (Seção 5.6) e heterogêneos (Seção 5.7), é importante comparar o desempenho da abordagem ao utilizar essas duas configurações de *pool*. Os resultados são apresentados para cada métrica avaliada: acurácia (Tabela 15), F-score (Tabela 16) e MCC (Tabela 17).

Avaliando o resultados apresentados nas Tabelas 15, 16 e 17, o PS-DES, configurado com *pool* heterogêneo, obteve melhores resultados de ranking para todas as métricas avaliadas, em comparação a todas as configurações de *pools* homogêneos. A análise estatística comprova que a configuração com *pool* heterogêneo supera as configurações com *pool* homogêneos em todas as três métricas avaliadas. Esses resultados sugerem a capacidade do PS-DES em avaliar as melhores técnicas SDE, especialmente quando confrontado com a diversidade gerada pelos classificadores heterogêneos.

Tabela 15 – Avaliação da acurácia média comparando a abordagem PS-DES utilizando um *pool* heterogêneo com três variações do PS-DES utilizando configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com *pool* heterogêneo com a configuração de *pools* homogêneos apresentada na coluna.

Base	NB	LR	Perceptron	Heterogêneo
appendicitis	0,882	0,862	0,835	0,865
banana	0,717	0,689	0,686	0,749
blood	0,768	0,787	0,784	0,787
bupa	0,598	0,670	0,670	0,667
cleveland	0,595	0,616	0,608	0,614
cmc	0,475	0,507	0,477	0,513
column_3C	0,809	0,847	0,850	0,845
dermatology	0,804	0,970	0,968	0,971
diabetes	0,753	0,768	0,756	0,766
ecoli-0_vs_1	0,792	0,964	0,972	0,972
german	0,740	0,746	0,723	0,744
glass1	0,638	0,677	0,685	0,682
glass6	0,954	0,951	0,937	0,947
haberman	0,744	0,737	0,732	0,733
hayes	0,701	0,669	0,663	0,685
heart	0,835	0,828	0,831	0,839
ilpd	0,603	0,711	0,711	0,687
led7digit	0,307	0,710	0,714	0,721
musk	0,765	0,773	0,777	0,797
new-thyroid1	0,939	0,959	0,953	0,956
pima	0,754	0,766	0,755	0,765
sonar	0,739	0,788	0,780	0,799
transfusion	0,770	0,789	0,781	0,785
vehicle	0,549	0,756	0,767	0,762
vehicle2	0,896	0,948	0,953	0,955
vowel	0,960	0,972	0,972	0,978
vowel0	0,959	0,972	0,970	0,978
wdbc	0,959	0,967	0,966	0,970
wine	0,968	0,970	0,971	0,975
yeast-2_vs_4	0,890	0,951	0,945	0,945
ranking	3,43	2,10	2,70	1,77
p-value	0,000	0,015	0,000	-

Fonte: Elaborada pelo autor (2024)

Tabela 16 – Avaliação do F-score média comparando a abordagem PS-DES utilizando um *pool* heterogêneo com três variações do PS-DES utilizando configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com *pool* heterogêneo com a configuração de *pools* homogêneo apresentada na coluna.

Base	NB	LR	Perceptron	Heterogêneo
appendicitis	0,818	0,726	0,647	0,758
banana	0,692	0,635	0,630	0,721
blood	0,612	0,585	0,549	0,610
bupa	0,590	0,642	0,625	0,641
cleveland	0,402	0,341	0,331	0,339
cmc	0,472	0,476	0,391	0,483
column_3C	0,757	0,801	0,801	0,798
dermatology	0,725	0,966	0,964	0,966
diabetes	0,721	0,728	0,697	0,726
ecoli-0_vs_1	0,760	0,960	0,969	0,968
german	0,685	0,640	0,532	0,645
glass1	0,581	0,552	0,562	0,557
glass6	0,888	0,888	0,845	0,876
haberman	0,595	0,557	0,522	0,554
hayes	0,719	0,677	0,670	0,701
heart	0,832	0,823	0,826	0,835
ilpd	0,597	0,547	0,510	0,570
led7digit	0,203	0,701	0,705	0,716
musk	0,760	0,764	0,768	0,791
new-thyroid1	0,864	0,916	0,900	0,908
pima	0,721	0,726	0,696	0,724
sonar	0,736	0,785	0,776	0,795
transfusion	0,615	0,590	0,541	0,613
vehicle	0,537	0,749	0,764	0,757
vehicle2	0,864	0,931	0,938	0,942
vowel	0,891	0,909	0,904	0,931
vowel0	0,888	0,909	0,899	0,932
wdbc	0,955	0,963	0,962	0,967
wine	0,968	0,970	0,971	0,975
yeast-2_vs_4	0,683	0,831	0,799	0,800
ranking	2,80	2,33	3,10	1,77
p-value	0,014	0,001	0,000	-

Fonte: Elaborada pelo autor (2024)

Tabela 17 – Avaliação do MCC médio comparando a abordagem PS-DES utilizando um *pool* heterogêneo com três variações do PS-DES utilizando configurações de *pool* homogêneos (NB, LR e Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha ranking apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com *pool* heterogêneo com a configuração de *pools* homogêneo apresentada na coluna.

Base	NB	LR	Perceptron	Heterogêneo
appendicitis	0,654	0,543	0,395	0,571
banana	0,441	0,421	0,419	0,528
blood	0,254	0,271	0,231	0,284
bupa	0,192	0,309	0,311	0,306
cleveland	0,358	0,349	0,322	0,345
cmc	0,223	0,224	0,157	0,237
column_3C	0,695	0,758	0,764	0,753
dermatology	0,771	0,963	0,960	0,964
diabetes	0,446	0,468	0,432	0,465
ecoli-0_vs_1	0,662	0,924	0,939	0,939
german	0,372	0,334	0,226	0,330
glass1	0,212	0,220	0,247	0,225
glass6	0,786	0,787	0,718	0,768
haberman	0,248	0,198	0,147	0,187
hayes	0,557	0,501	0,495	0,529
heart	0,669	0,656	0,665	0,677
ilpd	0,344	0,175	0,130	0,172
led7digit	0,263	0,681	0,684	0,692
musk	0,525	0,535	0,542	0,586
new-thyroid1	0,751	0,849	0,824	0,832
pima	0,447	0,465	0,430	0,461
sonar	0,482	0,580	0,563	0,600
transfusion	0,259	0,276	0,218	0,286
vehicle	0,422	0,678	0,692	0,685
vehicle2	0,732	0,864	0,877	0,884
vowel	0,791	0,828	0,826	0,867
vowel0	0,786	0,828	0,816	0,870
wdbc	0,913	0,926	0,925	0,934
wine	0,953	0,955	0,958	0,963
yeast-2_vs_4	0,405	0,695	0,651	0,653
ranking	3,03	2,23	3,00	1,73
p-value	0,002	0,008	0,000	-

Fonte: Elaborada pelo autor (2024)

5.10 ANÁLISE 5: AVALIAÇÃO DO PS-DES UTILIZANDO DIFERENTES FORMAS PARA CALCULAR O POTENCIAL.

Esta análise visa mostrar como o PS-DES se comporta ao utilizar quatro formas de avaliação do potencial dos *ensembles* (Etapa 2.3): acurácia (PS-DES-ACC), F-score (PS-DES-F), MCC (PS-DES-MCC) e HaO (PS-DES-HaO). Os resultados da análise estatística são apresentados na Tabela 18.

Tabela 18 – Resultado do ranking e do *p-valor* referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES com quatro formas de avaliação para o potencial (Etapa 2.3): acurácia (PS-DES-ACC), F-score (PS-DES-F), MCC (PS-DES-MCC) e HaO (PS-DES-HaO). São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	PS-DES-ACC	PS-DES-F	PS-DES-MCC	PS-DES-HaO
Acurácia	ranking	2,16	2,85	3,00	2,00
	p-valor	0,369	0,003	0,000	-
F-score	ranking	2,21	2,52	3,28	2,00
	p-valor	0,128	0,045	0,000	-
MCC	ranking	2,21	2,52	3,38	1,90
	p-valor	0,185	0,011	0,000	-

Fonte: Elaborada pelo autor (2024)

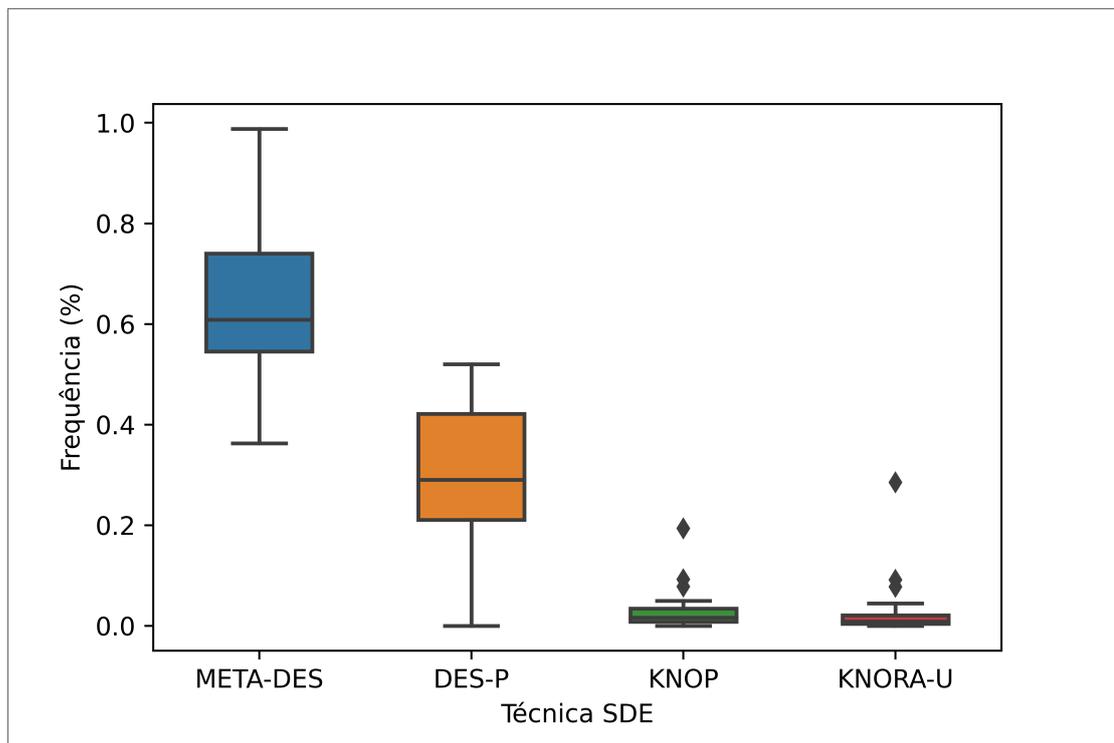
Avaliando a Tabela 18 é possível notar que o PS-DES-HaO como avaliação do potencial, tem o menor ranking (melhor desempenho) em todas as métricas. Já o PS-DES-MCC utilizando tem o maior ranking (pior desempenho). Analisando o resultado do teste estatístico, o PS-DES-HaO obteve melhores resultados em todas as métricas quando comparado ao PS-DES-F e PS-DES-MCC. Quando comparado ao PS-DES-ACC os resultados são equivalentes estatisticamente. Então, de modo geral, o HaO é sugerido como forma de avaliação dos *ensembles*. É interessante notar que os resultados não demonstraram correlação direta entre as métricas utilizadas para calcular o potencial dos *ensembles* e as utilizadas para avaliar as abordagens.

5.11 ANÁLISE 6: COMO FUNCIONA PS-DES NA PÓS-SELEÇÃO?

Esta análise visa mostrar como o PS-DES, utilizando o HaO como forma de pós-seleção, diversifica a pós-seleção de *ensembles* (Etapa 2.4) para diferentes situações. A Figura 7 apresenta o gráfico de *box-plot* com a frequência percentual da pós-seleção das técnicas SDE pelo

PS-DES para todas as bases de dados. A técnica META-DES é a mais escolhida, seguida da DES-P. Esses resultados não são surpreendentes, pois META-DES e DES-P são as técnicas SDE com o melhor desempenho individual. No entanto, dependendo do problema, KNORA-U e KNOP também são escolhidos. Essa diversificação das escolhas das técnicas é um dos fatores torna o PS-DES eficaz em múltiplos conjuntos de dados e melhora seu desempenho em comparação com as melhores técnicas individuais.

Figura 7 – Box-plot da frequência percentual de escolhas das técnicas SDE realizadas pelo PS-DES na Etapa 2.4, para todas as bases.

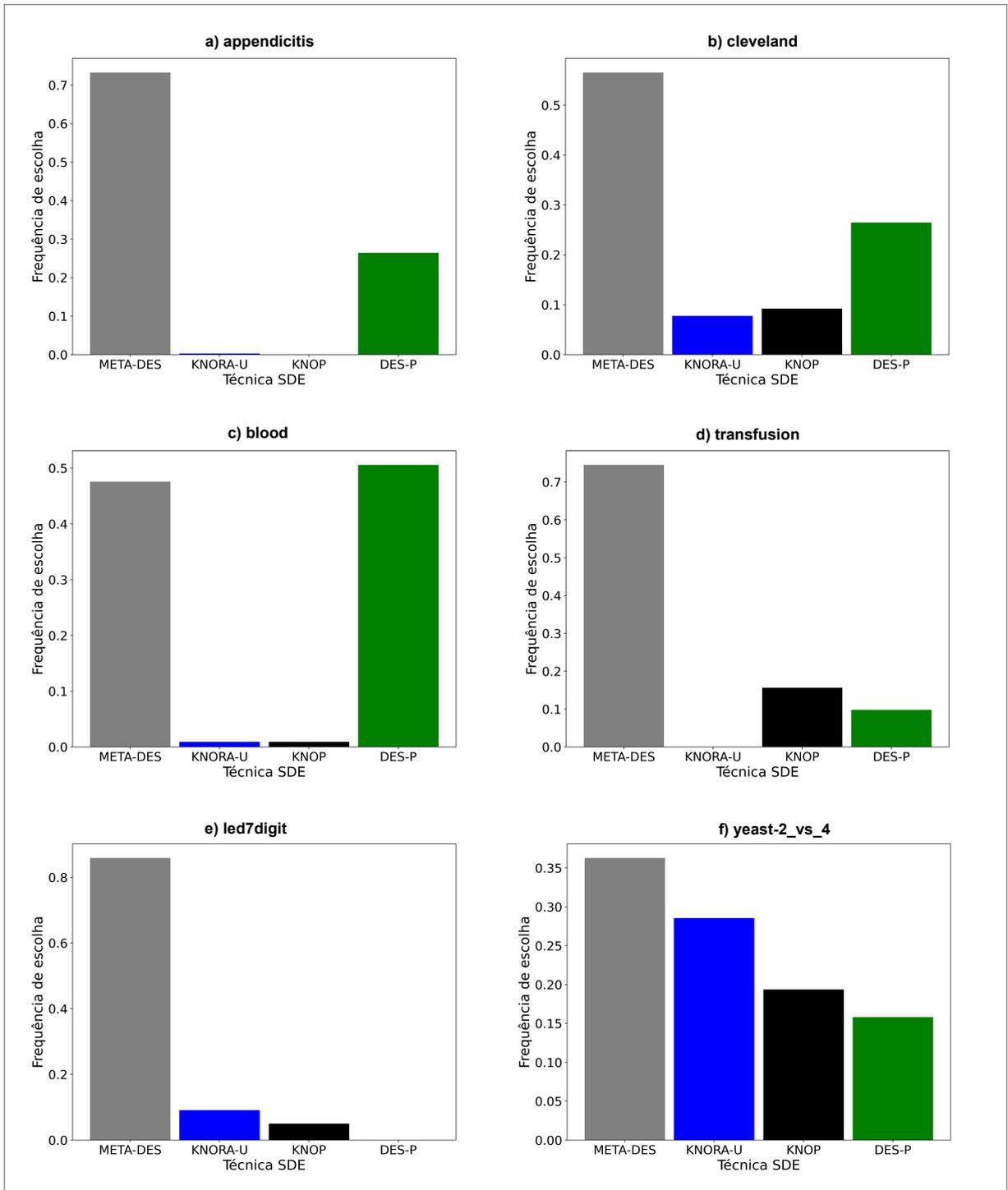


Fonte: Elaborada pelo autor(2024)

Objetivando apresentar alguns exemplos em bases, foi calculado quantas vezes (%) o PS-DES selecionou cada técnica individual SDE para seis base de dados (appendicitis, cleveland, blood, transfusion, led7digit e yeast-2_vs_4), e os resultados são mostrados na Figura 8. Os resultados da frequência de escolha para as outras bases de dados estão dispostos no Apêndice B.

A avaliação da Figura 8 mostra que o PS-DES seleciona diferentes técnicas SDE para diferentes situações. Isso pode explicar por que o PS-DES tem um desempenho melhor do que as técnicas individuais, como mostrados com os resultados dos experimentos deste capítulo.

Figura 8 – Frequência percentual das escolhas das técnicas SDE realizadas pelo PS-DES na Etapa 2.4, para as bases de dados: appendicitis, cleveland, blood, transfusion, led7digit e yeast-2_vs_4.



Fonte: Elaborada pelo autor(2024)

5.12 ANÁLISE 7: AVALIAÇÃO DO PS-DES-MVL

Inicialmente é avaliado o comportamento do PS-DES-MVL comparados as quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P), quando aplicadas, individualmente, à cinco funções de geração de *views*: t-SNE (Tabela 19), PCA (Tabela 20), *Kernel* Linear (Tabela 21), *Kernel* Polinomial (Tabela 22) e *Kernel* RBF (Tabela 23). São avaliadas três métricas (acurácia, F-score e MCC) e apresentados o ranking das abordagens, além do resultado do teste pareado de Wilcoxon (*p*-valor) ao comparar as técnicas SDE individuais ao PS-DES-MVL

Tabela 19 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o t-SNE como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-DES-MVL
Acurácia	ranking	3,03	3,62	3,26	2,60	2,48
	p-valor	0,021	0,003	0,023	0,470	-
F-score	ranking	3,86	4,52	2,31	2,45	1,86
	p-valor	0,000	0,000	0,478	0,006	-
MCC	ranking	3,83	4,31	2,48	2,41	1,97
	p-valor	0,000	0,000	0,152	0,061	-

Fonte: Elaborada pelo autor (2024)

Tabela 20 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o PCA como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-DES-MVL
Acurácia	ranking	2,88	3,32	3,37	2,63	2,80
	p-valor	0,293	0,208	0,038	0,786	-
F-score	ranking	3,63	4,03	2,50	2,80	2,03
	p-valor	0,000	0,000	0,086	0,008	-
MCC	ranking	3,10	3,47	3,23	2,77	2,43
	p-valor	0,017	0,009	0,041	0,136	-

Fonte: Elaborada pelo autor (2024)

Avaliando as Tabelas 19, 20, 21, 22, 23, nota-se que o PS-DES-MVL, de forma geral, obteve o menor ranking na maior parte da análise, além de ter superado as outras abordagem

Tabela 21 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o *kernel* linear como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-DES-MVL
Acurácia	ranking	2,36	2,62	3,59	3,67	2,76
	p-valor	0,915	0,853	0,003	0,025	-
F-score	ranking	3,47	3,38	2,38	3,64	2,14
	p-valor	0,000	0,000	0,444	0,000	-
MCC	ranking	2,85	2,93	3,10	3,67	2,45
	p-valor	0,043	0,066	0,061	0,000	-

Fonte: Elaborada pelo autor (2024)

Tabela 22 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o *kernel* polinomial como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-DES-MVL
Acurácia	ranking	2,57	2,81	3,57	3,45	2,60
	p-valor	0,863	0,832	0,001	0,115	-
F-score	ranking	3,71	3,71	2,35	3,38	1,86
	p-valor	0,000	0,000	0,173	0,000	-
MCC	ranking	3,02	2,98	3,28	3,45	2,28
	p-valor	0,058	0,047	0,001	0,001	-

Fonte: Elaborada pelo autor (2024)

Tabela 23 – Resultado do ranking e do *p*-valor referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL com quatro abordagens SDE (KNOP, KNORA-U, META-DES e DES-P) quando utilizado o *kernel* RBF como técnica de geração de *view*. São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	KNOP	KNORA-U	META-DES	DES-P	PS-DES-MVL
Acurácia	ranking	2,93	3,19	3,35	3,16	2,38
	p-valor	0,152	0,092	0,000	0,061	-
F-score	ranking	3,48	3,86	2,48	3,21	1,97
	p-valor	0,000	0,000	0,099	0,001	-
MCC	ranking	3,31	3,55	2,83	3,35	1,97
	p-valor	0,001	0,000	0,006	0,000	-

Fonte: Elaborada pelo autor (2024)

em 37 das 60 comparações, avaliando o p-valor. Contudo, não existe uma *view* na qual o PS-DES-MVL obteve melhor em todas as métricas, quando comparado as abordagens SDE. Exemplificando: o PS-DES-MVL obteve resultados superiores ao DES-P, em todas as métricas, quando utilizado o *kernel* linear como função de geração de *views*. Contudo obteve resultados similares (em acurácia e MCC) quando avaliado o PCA como função de geração de *views*. É possível sugerir que a técnica de geração de *view* influencia o comportamento dos algoritmos individuais sendo um caminho viável para adicionar diversidade a um SMC, e, a depender da *view* utilizada, a abordagem proposta torna-se uma possível candidata para melhorar o desempenho dos modelos SDE individuais.

Na próxima análise é avaliado o comportamento do PS-DES-MVL quando configurados com cinco funções de geração de *views* individuais e combinadas. São avaliadas três métricas (acurácia, F-score e MCC) e apresentados o ranking das abordagens, além do resultado do teste pareado de Wilcoxon (p-valor) ao comparar as configurações de *views* individuais e combinadas. Os resultados são apresentados na Tabela 24.

Tabela 24 – Resultado do ranking e do *p-valor* referente à aplicação do teste de Wilcoxon comparando o desempenho do PS-DES-MVL quando utilizado cinco funções de representação de dados: t-SNE, PCA, *Kernel* Linear (K-Linear), *Kernel* Polinomial (K-Polinomial) e *Kernel* RBF (K-RBF), de forma individual e combinadas (MVL). São avaliadas três métricas de desempenho (Acurácia, F-score e MCC). Todas as abordagens utilizam um *pool* heterogêneo com 100 classificadores NB, 100 classificadores LR e 100 Perceptron.

Métrica	Critério	t-SNE	PCA	K-Linear	K-Polinomial	K-RBF	PS-DES-MVL
Acurácia	ranking	4,69	2,72	3,26	3,67	2,79	3,86
	p-valor	0,019	0,999	0,951	0,832	0,946	-
F-score	ranking	4,79	2,83	2,41	2,93	2,93	5,10
	p-valor	0,745	1,000	1,000	1,000	0,999	-
MCC	ranking	4,83	2,59	2,76	3,17	2,97	4,69
	p-valor	0,313	1,000	1,000	1,000	0,992	-

Fonte: Elaborada pelo autor (2024)

Além da avaliação comparativa do desempenho da abordagem PS-DES-MVL também foi avaliado a diversidade, comparando o PS-DES-MVL com o PS-DES para as métricas DF e DIS. O resultado estatístico, utilizando o DIS, sugere que a abordagem PS-DES-MVL é mais diversa do que a PS-DES (p-valor = 0.000). Quando utilizado o DF para calcular a diversidade, a análise estatística (p-valor = 0.999) sugere que o PS-DES-MVL é menos diverso do que o PS-DES. Com isso, mesmo que fosse esperado a diversidade da abordagem MVL, não foi possível confirmar perante a abordagem tradicional.

Os resultados da Tabela 24 mostram que o PS-DES-MVL, utilizando a configuração com as *views* combinadas, não obteve resultados promissores na comparação com as *views* individuais. Apenas quando comparado ao t-SNE e na acurácia, a abordagem MVL obteve resultados superiores. Uma possível explicação para os resultados negativos foi a escolha do kDN como métrica de seleção de *view*. Como a abordagem teve resultados promissores quando comparados a técnicas SDE individuais (Tabelas 19, 20, 21, 22, 23), uma possível explicação é que o kDN não seja a métrica mais adequada para definir a melhor *view*.

6 CONCLUSÃO

Nesta tese, foram investigadas duas novidades relacionados às técnicas de SDE: (i) o desenvolvimento de uma estratégia eficaz para selecionar um ensemble dado um conjunto de *ensembles* gerados por diferentes técnicas SDE e (ii) como Sistema de múltiplos classificadores (SMC) comportam-se ao incorporar múltiplas técnicas *Multi-view learning*.

Para abordar a primeira questão de pesquisa, foi proposta uma nova abordagem, denominada PS-DES, que avalia e seleciona *ensembles* gerados à partir da fase de seleção de um conjunto de técnicas SDE. O PS-DES baseia-se na suposição de que diferentes critérios de seleção podem resultar em diversos *ensembles* selecionados, e os critérios ideais para selecionar um conjunto podem variar dependendo da amostra avaliada. O PS-DES opera como uma estratégia de pós-seleção que avalia um conjunto de técnicas individuais SDE e escolhe aquele com o maior HaO, objetivando melhorar o desempenho quando das técnicas individuais SDE. O HaO é baseado no Oráculo, contudo é capaz de analisar não apenas a existência de um único classificador competente, mas também a capacidade de uma técnica SDE em encontrar um número significativo desses classificadores.

Os resultados da avaliação experimental mostraram que o PS-DES, em conjunto com o HaO, é capaz de avaliar *ensembles* de forma mais eficaz do que as métricas tradicionais de avaliação de classificador-a-classificador. Essa afirmação é confirmada pelos experimentos e análises estatísticas apresentados durante a tese, em uma variedade de bases de dados para problemas de classificação. Os melhores resultados obtidos concentram-se nas configurações com *pools* heterogêneos, que, assim como mostrado em um dos experimentos deste trabalho, são mais diversos do que os *pools* homogêneos, indicando que o PS-DES se adapta melhor a este tipo de configuração. Além disso, foi mostrado que o PS-DES é capaz de selecionar diferentes técnicas SDE para diferentes situações, o que pode explicar os melhores resultados, avaliados estatisticamente, quando comparados a técnicas individuais SDE.

Para a segunda questão de pesquisa, foi proposta uma atualização o PS-DES para incorporar conceitos de MVL, especificamente a geração de múltiplas *views* objetivando a geração de *pools* mais diversos, denominado PS-DES-MVL. Os resultados mostraram que a técnica de geração de *view* influencia o comportamento dos algoritmos individuais, sendo um caminho viável para adicionar diversidade a um SMC. A depender da *view* utilizada, a abordagem proposta torna-se uma possível candidata para melhorar o desempenho dos modelos SDE in-

individuais. Contudo, ao combinar as *views* os resultados não foram promissores, sugerindo que a forma de seleção de *view*, utilizando IH, precisa de mais maturação.

6.1 TRABALHOS FUTUROS

Esta tese deixa algumas lacunas a serem preenchidas em trabalhos futuros. Inicialmente, é proposto uma análise matemática do HaO. Além disso, pode-se citar uma avaliação extensiva de novos classificadores-base para a formação dos *pools*. O PS-DES foi avaliado usando três classificadores diferentes, logo ainda existe espaço para avaliar seu desempenho usando classificadores-base mais robustos (*e.g.*, SVM, MLP, etc.). A exploração de novas técnicas SDE pode enriquecer a avaliação deste trabalho, o que não pôde ser atingido devido a problemas com o tempo, além da reprodutibilidade das outras técnicas discutidas na literatura. A avaliação PS-DES proposta em bases de dados maiores também é um ponto a ser avaliado posteriormente.

Sobre o PS-DES-MVL, é importante avaliar outras técnicas de geração de *views*, com características distintas das utilizadas nesta tese. O objetivo é encontrar novas formas de gerar diversidades ao sistema. Como o kDN não obteve resultados satisfatórios para a seleção da melhor *view*, sugere-se avaliar novas abordagens para realizar esta seleção. Finalmente, é interessante avaliar a abordagem em aplicações nas quais as *views* já estão definidas como MVL, sem a necessidade de aplicar técnicas de geração, ou seja, de forma artificial.

6.2 TRABALHOS PUBLICADOS

Os trabalhos mostrados a seguir foram resultados da pesquisa realizada durante o doutoramento:

- **A post-selection algorithm for improving dynamic ensemble selection methods** (CORDEIRO; CAVALCANTI; CRUZ, 2023a)
- **Dynamic Ensemble Algorithm Post-Selection Using Hardness-Aware Oracle** (CORDEIRO; CAVALCANTI; CRUZ, 2023b)

REFERÊNCIAS

- ACHAKZAI, M. A. K.; PENG, J. Detecting financial statement fraud using dynamic ensemble machine learning. *International Review of Financial Analysis*, v. 89, p. 102827, 2023. ISSN 1057-5219.
- AMORIM, L. B. de; CAVALCANTI, G. D.; CRUZ, R. M. The choice of scaling technique matters for classification performance. *Applied Soft Computing*, Elsevier, v. 133, p. 109924, 2023.
- BATCHU, R. K.; SEETHA, H. An integrated approach explaining the detection of distributed denial of service attacks. *Computer Networks*, v. 216, p. 109269, 2022. ISSN 1389-1286.
- BENAVOLI, A.; CORANI, G.; MANGILI, F. Should we really use post-hoc tests based on mean-ranks? *The Journal of Machine Learning Research*, JMLR. org, v. 17, n. 1, p. 152–161, 2016.
- BIAN, S.; WANG, W. On diversity and accuracy of homogeneous and heterogeneous ensembles. *International Journal of Hybrid Intelligent Systems*, IOS Press, v. 4, n. 2, p. 103–128, 2007.
- BOCK, K. W. D.; COUSSEMENT, K.; LESSMANN, S. Cost-sensitive business failure prediction when misclassification costs are uncertain: A heterogeneous ensemble selection approach. *European Journal of Operational Research*, Elsevier, 2020.
- BREIMAN, L. Bagging predictors. *Machine learning*, Springer, v. 24, n. 2, p. 123–140, 1996.
- CAVALIN, P. R.; SABOURIN, R.; SUEN, C. Y. Dynamic selection approaches for multiple classifier systems. *Neural Computing and Applications*, Springer, v. 22, p. 673–688, 2013.
- CHONGOMWERU, H.; KASEM, A. A novel ensemble method for classification in imbalanced datasets using split balancing technique based on instance hardness (sbal_ih). *Neural Computing and Applications*, Springer, v. 33, n. 17, p. 11233–11254, 2021.
- CORDEIRO, P. R.; CAVALCANTI, G. D.; CRUZ, R. M. A post-selection algorithm for improving dynamic ensemble selection methods. In: IEEE. *2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. [S.l.], 2023. p. 1142–1147.
- CORDEIRO, P. R. G.; CAVALCANTI, G. D. C.; CRUZ, R. M. O. Dynamic ensemble algorithm post-selection using hardness-aware oracle. *IEEE Access*, v. 11, p. 86056–86070, 2023.
- COSTA, V. S.; FARIAS, A. D. S.; BEDREGAL, B.; SANTIAGO, R. H.; CANUTO, A. M. d. P. Combining multiple algorithms in classifier ensembles using generalized mixture functions. *Neurocomputing*, v. 313, p. 402–414, 2018.
- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, Elsevier B.V., v. 41, p. 195–216, 2018.
- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D. Prototype selection for dynamic classifier and ensemble selection. *Neural Computing and Applications*, Springer, v. 29, p. 447–457, 2018.

- CRUZ, R. M.; SABOURIN, R.; CAVALCANTI, G. D.; REN, T. I. Meta-des: A dynamic ensemble selection framework using meta-learning. *Pattern Recognition*, Elsevier BV, v. 48, n. 5, p. 1925–1935, 2015.
- CRUZ, R. M.; SOUZA, M. A.; SABOURIN, R.; CAVALCANTI, G. D. Dynamic ensemble selection and data preprocessing for multi-class imbalance learning. *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific, v. 33, n. 11, p. 1940009, 2019.
- CRUZ, R. M. O.; HAFEMANN, L. G.; SABOURIN, R.; CAVALCANTI, G. D. C. Deslib: A dynamic ensemble selection library in python. *Journal of Machine Learning Research*, v. 21, n. 8, p. 1–5, 2020.
- DAVTALAB, R.; CRUZ, R. M.; SABOURIN, R. Dynamic ensemble selection using fuzzy hyperboxes. In: IEEE. *2022 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2022. p. 1–9.
- DAVTALAB, R.; CRUZ, R. M.; SABOURIN, R. A scalable dynamic ensemble selection using fuzzy hyperboxes. *Information Fusion*, v. 102, p. 102036, 2024. ISSN 1566-2535.
- DIETTERICH, T. G. An experimental comparison of three methods for constructing ensembles of decision trees: Bagging, boosting, and randomization. *Machine learning*, Springer, v. 40, p. 139–157, 2000.
- DO, D. T.; NGUYEN, T. T.; NGUYEN, T. T.; LUONG, A. V.; LIEW, A. W.-C.; MCCALL, J. Confidence in prediction: An approach for dynamic weighted ensemble. In: SPRINGER. *Intelligent Information and Database Systems: 12th Asian Conference, ACIIDS 2020, Phuket, Thailand, March 23–26, 2020, Proceedings, Part I 12*. [S.l.], 2020. p. 358–370.
- DUA, D.; GRAFF, C. *UCI Machine Learning Repository*. 2020. Disponível em: <<http://archive.ics.uci.edu/ml>>.
- DUIN, R. The combining classifier: to train or not to train? In: *2002 International Conference on Pattern Recognition*. [S.l.: s.n.], 2002. v. 2, p. 765–770 vol.2.
- EL-SAPPAGH, S.; ALONSO-MORAL, J. M.; ABUHMED, T.; ALI, F.; BUGARÍN-DIZ, A. Trustworthy artificial intelligence in alzheimer’s disease: state of the art, opportunities, and challenges. *Artificial Intelligence Review*, Springer, v. 56, n. 10, p. 11149–11296, 2023.
- ELMI, J.; EFTEKHARI, M. Multi-layer selector (mls): Dynamic selection based on filtering some competence measures. *Applied Soft Computing*, v. 104, p. 107257, 2021.
- FILHO, J. A. S. L.; CANUTO, A. M.; SANTIAGO, R. H. N. Investigating the impact of selection criteria in dynamic ensemble selection methods. *Expert Systems with Applications*, Elsevier, v. 106, p. 141–153, 2018.
- FREUND, Y.; SCHAPIRE, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, v. 55, n. 1, p. 119–139, 1997. ISSN 0022-0000.
- GUO, H.; LIU, H.; LI, R.; WU, C.; GUO, Y.; XU, M. Margin & diversity based ordering ensemble pruning. *Neurocomputing*, v. 275, p. 237–246, 2018.

- HAQUE, M. N.; NOMAN, N.; BERRETTA, R.; MOSCATO, P. Heterogeneous ensemble combination search using genetic algorithm for class imbalanced data classification. *PloS One*, Public Library of Science, v. 11, n. 1, 2016.
- JADHAV, A.; SHANDILYA, S. K.; IZONIN, I.; GREGUS, M. Effective software effort estimation enabling digital transformation. *IEEE Access*, p. 1–1, 2023.
- JR, A. S. B.; SABOURIN, R.; OLIVEIRA, L. E. Dynamic selection of classifiers—a comprehensive review. *Pattern recognition*, Elsevier, v. 47, n. 11, p. 3665–3680, 2014.
- JUNIOR, L. M.; NARDINI, F. M.; RENSO, C.; TRANI, R.; MACEDO, J. A. A novel approach to define the local region of dynamic selection techniques in imbalanced credit scoring problems. *Expert Systems with Applications*, Elsevier, v. 152, p. 113351, 2020.
- KADKHODAEI, H. R.; MOGHADAM, A. M. E.; DEGHAN, M. Hboost: A heterogeneous ensemble classifier based on the boosting method and entropy measurement. *Expert Systems with Applications*, Elsevier, p. 113482, 2020.
- KITTLER, J. Combining classifiers: A theoretical framework. *Pattern analysis and Applications*, Springer, v. 1, p. 18–27, 1998.
- KITTLER, J.; HATEF, M.; DUIN, R. P.; MATAS, J. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 20, n. 3, p. 226–239, 1998.
- KO, A. H.; SABOURIN, R.; JR, A. S. B. From dynamic classifier selection to dynamic ensemble selection. *Pattern recognition*, Elsevier, v. 41, n. 5, p. 1718–1731, 2008.
- KUMAR, V.; MINZ, S. Multi-view ensemble learning: an optimal feature set partitioning for high-dimensional data classification. *Knowledge and Information Systems*, Springer, v. 49, n. 1, p. 1–59, 2016.
- KUNCHEVA, L. I. Clustering-and-selection model for classifier combination. In: IEEE. *KES'2000. Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies. Proceedings (Cat. No. 00TH8516)*. [S.l.], 2000. v. 1, p. 185–188.
- KUNCHEVA, L. I. A theoretical study on six classifier fusion strategies. *IEEE Transactions on pattern analysis and machine intelligence*, IEEE, v. 24, n. 2, p. 281–286, 2002.
- KUNCHEVA, L. I. *Combining Pattern Classifiers: Methods and Algorithms: Second Edition*. [S.l.]: Wiley Publishing, 2014. 1–357 p. ISBN 9781118914564.
- KUNCHEVA, L. I.; WHITAKER, C. J. Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine learning*, v. 51, n. 2, p. 181, 2003.
- LARGE, J.; LINES, J.; BAGNALL, A. The heterogeneous ensembles of standard classification algorithms (hesca): the whole is greater than the sum of its parts. *arXiv preprint arXiv:1710.09220*, 2017.
- LE, T.-B.; KIM, S.-W. On measuring confidence levels using multiple views of feature set for useful unlabeled data selection. *Neurocomputing*, Elsevier, v. 173, p. 1589–1601, 2016.

- LE, T.-B.; KIM, S.-W. Transformation based methods extracting multiple views for co-training algorithms. In: *Proceedings on the International Conference on Artificial Intelligence (ICAI)*. [S.l.: s.n.], 2017. p. 97–103.
- LI, Y.; YANG, M.; ZHANG, Z. A survey of multi-view representation learning. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, v. 31, n. 10, p. 1863–1883, 2018.
- LI, Y.; YANG, M.; ZHANG, Z. M. A Survey of Multi-View Representation Learning. *IEEE Transactions on Knowledge and Data Engineering*, IEEE, PP, n. 8, p. 1, 2018.
- LU, H.; SU, H.; HU, J.; DU, Q. Dynamic ensemble learning with multi-view kernel collaborative subspace clustering for hyperspectral image classification. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, IEEE, v. 15, p. 2681–2695, 2022.
- MADDINENI, R. D.; CH, D. Consistent interpretation of ensemble classifiers in trojan-horse detection. *IEEE Access*, v. 11, p. 70930–70946, 2023.
- MASOUDNIA, S.; EBRAHIMPOUR, R. Mixture of experts: a literature survey. *Artificial Intelligence Review*, Springer, v. 42, p. 275–293, 2014.
- NANNI, L.; BRAHNAM, S.; GHIDONI, S.; LUMINI, A. Toward a general-purpose heterogeneous ensemble for pattern classification. *Computational Intelligence and Neuroscience*, Hindawi, v. 2015, 2015.
- NGUYEN, T. T.; DANG, M. T.; LIEW, A. W.; BEZDEK, J. C. A weighted multiple classifier framework based on random projection. *Information Sciences*, v. 490, p. 36–58, 2019.
- NGUYEN, T. T.; LUONG, A. V.; DANG, M. T.; LIEW, A. W. C.; MCCALL, J. Ensemble Selection based on Classifier Prediction Confidence. *Pattern Recognition*, Elsevier Ltd, v. 100, p. 107104, 2020.
- NGUYEN, T. T.; NGUYEN, M. P.; PHAM, X. C.; LIEW, A. W.-C.; PEDRYCZ, W. Combining heterogeneous classifiers via granular prototypes. *Applied Soft Computing*, Elsevier, v. 73, p. 795–815, 2018.
- OLIVEIRA, D. V.; CAVALCANTI, G. D.; SABOURIN, R. Online pruning of base classifiers for dynamic ensemble selection. *Pattern Recognition*, v. 72, p. 44–58, 2017. ISSN 0031-3203.
- OSTVAR, N.; MOGHADAM, A. M. E. Hdec: A heterogeneous dynamic ensemble classifier for binary datasets. *Computational Intelligence and Neuroscience*, Hindawi, v. 2020, 2020.
- PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.
- PEKALSKA, E.; DUIN, R. P. On combining dissimilarity representations. In: SPRINGER. *International Workshop on Multiple Classifier Systems*. [S.l.], 2001. p. 359–368.
- RODRIGUES, W. M.; WALMSLEY, F. N.; CAVALCANTI, G. D. C.; CRUZ, R. M. O. Security relevant methods of android's api classification: A machine learning empirical evaluation. *IEEE Transactions on Computers*, p. 1–13, 2023.

- SMITH, M. R.; MARTINEZ, T.; GIRAUD-CARRIER, C. An instance level analysis of data complexity. *Machine learning*, Springer, v. 95, p. 225–256, 2014.
- SOUZA, M. A.; CAVALCANTI, G. D. C.; CRUZ, R. M. O.; SABOURIN, R. On the characterization of the oracle for dynamic classifier selection. In: *International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2017. p. 332–339.
- SOUZA, M. A.; SABOURIN, R.; CAVALCANTI, G. D.; CRUZ, R. M. Local overlap reduction procedure for dynamic ensemble selection. In: IEEE. *2022 International Joint Conference on Neural Networks (IJCNN)*. [S.l.], 2022. p. 1–9.
- SOUZA, M. A.; SABOURIN, R.; CAVALCANTI, G. D.; CRUZ, R. M. Olp++: An online local classifier for high dimensional data. *Information Fusion*, Elsevier, v. 90, p. 120–137, 2023.
- SUN, S.; MAO, L.; DONG, Z.; WU, L. *Multiview Machine Learning*. [S.l.]: Springer, 2019.
- SUN, S.; MAO, L.; DONG, Z.; WU, L. View construction. In: *Multiview Machine Learning*. Singapore: Springer Singapore, 2019. p. 139–149. ISBN 9789811330285.
- SWAMINATHAN, B.; PALANI, S.; VAIRAVASUNDARAM, S. Meta learning-based dynamic ensemble model for crop selection. *Applied Artificial Intelligence*, Taylor & Francis, v. 36, n. 1, p. 2145646, 2022.
- TSOUMAKAS, G.; ANGELIS, L.; VLAHAVAS, I. Selective fusion of heterogeneous classifiers. *Intelligent Data Analysis*, IOS Press, v. 9, n. 6, p. 511–525, 2005.
- VITÓRIO, J. G. B.; JUNIOR, A. S. B.; G., C. Y. M.; SILLA, C. N. Multi-modal music mood classification with dynamic classifier selection. In: *2023 30th International Conference on Systems, Signals and Image Processing (IWSSIP)*. [S.l.: s.n.], 2023. p. 1–5.
- WALMSLEY, F. N.; CAVALCANTI, G. D.; OLIVEIRA, D. V.; CRUZ, R. M.; SABOURIN, R. An ensemble generation method based on instance hardness. In: IEEE. *2018 international joint conference on neural networks (IJCNN)*. [S.l.], 2018. p. 1–8.
- WANG, L.; MO, T.; WANG, X.; CHEN, W.; HE, Q.; LI, X.; ZHANG, S.; YANG, R.; WU, J.; GU, X. et al. A hierarchical fusion framework to integrate homogeneous and heterogeneous classifiers for medical decision-making. *Knowledge-Based Systems*, Elsevier, v. 212, p. 106517, 2021.
- WANG, Y.; ZHANG, J.; YAN, W. An enhanced dynamic ensemble selection classifier for imbalance classification with application to china corporation bond default prediction. *IEEE Access*, v. 11, p. 32082–32094, 2023.
- WANG, Z.; CHEN, S.; SUN, T. Multik-mhks: a novel multiple kernel learning algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, IEEE, v. 30, n. 2, p. 348–353, 2007.
- WANG, Z.; ZHAO, S.; LI, Z.; CHEN, H.; LI, C.; SHEN, Y. Ensemble selection with joint spectral clustering and structural sparsity. *Pattern Recognition*, v. 119, p. 108061, 2021.
- WANG, Z.; ZHU, Y.; CHEN, Z.; ZHANG, J.; DU, W. Multi-view learning with fisher kernel and bi-bagging for imbalanced problem. *Applied Intelligence*, Springer, v. 49, n. 8, p. 3109–3122, 2019.

WOLOSZYNSKI, T.; KURZYNSKI, M. A probabilistic model of classifier competence for dynamic ensemble selection. *Pattern Recognition*, v. 44, n. 10, p. 2656–2668, 2011.

WOLOSZYNSKI, T.; KURZYNSKI, M.; PODSIADLO, P.; STACHOWIAK, G. W. A measure of competence based on random classification for dynamic ensemble selection. *Information Fusion*, v. 13, n. 3, p. 207–213, 2012.

XIA, Y.; ZHAO, J.; HE, L.; LI, Y.; NIU, M. A novel tree-based dynamic heterogeneous ensemble method for credit scoring. *Expert Systems with Applications*, Elsevier, v. 159, p. 113615, 2020.

XU, C.; TAO, D.; XU, C. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*, 2013.

ZHAO, J.; XIE, X.; XU, X.; SUN, S. Multi-view learning overview: Recent progress and new challenges. *Information Fusion*, Elsevier, v. 38, p. 43–54, 2017.

ZHOU, Z.-H. *Ensemble Methods: Foundations and Algorithms*. 1st. ed. [S.l.]: Chapman and Hall/CRC, 2012. ISBN 1439830037.

ZHU, X.; LI, J.; REN, J.; WANG, J.; WANG, G. Dynamic ensemble learning for multi-label classification. *Information Sciences*, Elsevier, v. 623, p. 94–111, 2023.

ZYBLEWSKI, P.; SABOURIN, R.; WOŹNIAK, M. Preprocessed dynamic classifier ensemble selection for highly imbalanced drifted data streams. *Information Fusion*, v. 66, p. 138–154, 2021.

APÊNDICE A – RESULTADOS DA AVALIAÇÃO DO PS-DES UTILIZANDO POOLS HOMOGÊNEOS

Tabela 25 – Avaliação da taxa de acerto média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (NB). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento médio para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,881	0,879	0,883	0,879	0,878	0,882
banana	0,670	0,642	0,724	0,715	0,676	0,717
blood	0,764	0,764	0,762	0,769	0,763	0,768
bupa	0,579	0,574	0,578	0,597	0,595	0,598
cleveland	0,594	0,598	0,602	0,595	0,594	0,595
cmc	0,470	0,470	0,471	0,480	0,474	0,475
column_3C	0,810	0,807	0,810	0,808	0,806	0,809
dermatology	0,806	0,806	0,807	0,804	0,807	0,804
diabetes	0,752	0,754	0,750	0,754	0,752	0,753
ecoli-0_vs_1	0,791	0,791	0,790	0,791	0,791	0,792
german	0,737	0,736	0,738	0,738	0,738	0,740
glass1	0,587	0,596	0,634	0,641	0,618	0,638
glass6	0,954	0,952	0,954	0,954	0,954	0,954
haberman	0,745	0,746	0,736	0,749	0,740	0,744
hayes	0,658	0,663	0,687	0,680	0,664	0,701
heart	0,830	0,827	0,833	0,831	0,830	0,835
ilpd	0,587	0,582	0,600	0,596	0,587	0,603
led7digit	0,338	0,232	0,241	0,236	0,250	0,307
musk	0,746	0,746	0,766	0,763	0,753	0,765
new-thyroid1	0,941	0,938	0,940	0,943	0,938	0,939
pima	0,753	0,753	0,748	0,755	0,753	0,754
sonar	0,690	0,692	0,734	0,722	0,715	0,739
transfusion	0,764	0,764	0,764	0,770	0,770	0,770
vehicle	0,522	0,494	0,560	0,520	0,524	0,549
vehicle2	0,845	0,828	0,906	0,885	0,868	0,896
vowel	0,943	0,941	0,965	0,947	0,949	0,960
vowel0	0,944	0,940	0,963	0,946	0,949	0,959
wdbc	0,960	0,960	0,961	0,960	0,960	0,959
wine	0,968	0,968	0,969	0,968	0,967	0,968
yeast-2_vs_4	0,886	0,874	0,886	0,866	0,885	0,890
ranking	4,12	4,75	2,70	3,10	3,95	2,38
p-value	0,000	0,000	0,149	0,012	0,000	-

Tabela 26 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (NB). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,818	0,816	0,818	0,815	0,814	0,818
banana	0,636	0,602	0,700	0,688	0,642	0,692
blood	0,577	0,577	0,588	0,610	0,580	0,612
bupa	0,568	0,561	0,569	0,585	0,587	0,590
cleveland	0,400	0,409	0,412	0,409	0,407	0,402
cmc	0,469	0,469	0,468	0,477	0,472	0,472
column_3C	0,759	0,754	0,759	0,755	0,752	0,757
dermatology	0,728	0,730	0,729	0,726	0,730	0,725
diabetes	0,720	0,723	0,718	0,721	0,720	0,721
ecoli-0_vs_1	0,758	0,758	0,758	0,758	0,759	0,760
german	0,681	0,681	0,684	0,681	0,683	0,685
glass1	0,541	0,548	0,586	0,590	0,572	0,581
glass6	0,896	0,890	0,887	0,898	0,895	0,888
haberman	0,581	0,584	0,565	0,591	0,574	0,595
hayes	0,675	0,679	0,706	0,698	0,681	0,719
heart	0,828	0,824	0,831	0,829	0,828	0,832
ilpd	0,584	0,579	0,595	0,593	0,584	0,597
led7digit	0,231	0,138	0,167	0,146	0,169	0,203
musk	0,741	0,741	0,762	0,757	0,749	0,760
new-thyroid1	0,875	0,864	0,866	0,879	0,861	0,864
pima	0,721	0,721	0,715	0,722	0,721	0,721
sonar	0,689	0,691	0,732	0,720	0,714	0,736
transfusion	0,577	0,578	0,591	0,612	0,595	0,615
vehicle	0,500	0,469	0,549	0,500	0,508	0,537
vehicle2	0,790	0,770	0,876	0,848	0,825	0,864
vowel	0,851	0,847	0,903	0,860	0,865	0,891
vowel0	0,853	0,846	0,898	0,858	0,864	0,888
wdbc	0,956	0,957	0,957	0,956	0,956	0,955
wine	0,968	0,968	0,969	0,968	0,968	0,968
yeast-2_vs_4	0,582	0,596	0,594	0,675	0,623	0,683
ranking	4,12	4,75	2,70	3,10	3,95	2,38
p-value	0,000	0,000	0,149	0,012	0,000	-

Tabela 27 – Avaliação do MCC comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (NB). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,653	0,650	0,653	0,650	0,646	0,654
banana	0,339	0,275	0,459	0,438	0,355	0,441
blood	0,206	0,206	0,214	0,254	0,204	0,254
bupa	0,184	0,179	0,146	0,180	0,192	0,192
cleveland	0,357	0,366	0,369	0,363	0,361	0,358
cmc	0,224	0,224	0,216	0,232	0,227	0,223
column_3C	0,696	0,692	0,697	0,694	0,690	0,695
dermatology	0,773	0,774	0,775	0,770	0,775	0,771
diabetes	0,444	0,449	0,439	0,447	0,443	0,446
ecoli-0_vs_1	0,659	0,659	0,658	0,659	0,660	0,662
german	0,365	0,363	0,371	0,365	0,368	0,372
glass1	0,159	0,180	0,212	0,228	0,202	0,212
glass6	0,801	0,791	0,788	0,803	0,799	0,786
haberman	0,236	0,241	0,202	0,257	0,221	0,248
hayes	0,488	0,496	0,532	0,525	0,498	0,557
heart	0,660	0,653	0,666	0,661	0,660	0,669
ilpd	0,352	0,350	0,344	0,358	0,344	0,344
led7digit	0,304	0,167	0,178	0,174	0,190	0,263
musk	0,486	0,486	0,529	0,520	0,502	0,525
new-thyroid1	0,775	0,750	0,755	0,778	0,748	0,751
pima	0,446	0,446	0,435	0,449	0,446	0,447
sonar	0,396	0,401	0,476	0,451	0,441	0,482
transfusion	0,204	0,206	0,222	0,256	0,234	0,259
vehicle	0,393	0,358	0,436	0,389	0,394	0,422
vehicle2	0,589	0,546	0,756	0,701	0,655	0,732
vowel	0,717	0,711	0,814	0,732	0,743	0,791
vowel0	0,720	0,709	0,803	0,729	0,741	0,786
wdbc	0,915	0,916	0,917	0,915	0,914	0,913
wine	0,953	0,953	0,954	0,953	0,952	0,953
yeast-2_vs_4	0,228	0,246	0,244	0,387	0,307	0,405
ranking	3,58	3,78	2,63	2,67	3,90	2,33
p-value	0,002	0,000	0,154	0,038	0,000	-

Tabela 28 – Avaliação da taxa de acerto média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (LR). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,833	0,830	0,850	0,850	0,839	0,862
banana	0,649	0,649	0,688	0,692	0,655	0,689
blood	0,781	0,782	0,786	0,788	0,783	0,787
bupa	0,674	0,672	0,667	0,680	0,676	0,670
cleveland	0,622	0,616	0,615	0,619	0,614	0,616
cmc	0,504	0,503	0,509	0,507	0,506	0,507
column_3C	0,848	0,849	0,846	0,848	0,847	0,847
dermatology	0,969	0,969	0,970	0,967	0,969	0,970
diabetes	0,771	0,770	0,770	0,769	0,770	0,768
ecoli-0_vs_1	0,962	0,963	0,964	0,962	0,961	0,964
german	0,742	0,742	0,746	0,744	0,742	0,746
glass1	0,638	0,641	0,678	0,690	0,658	0,677
glass6	0,940	0,939	0,950	0,948	0,947	0,951
haberman	0,730	0,727	0,729	0,735	0,727	0,737
hayes	0,588	0,565	0,671	0,606	0,610	0,669
heart	0,826	0,827	0,828	0,827	0,827	0,828
ilpd	0,709	0,711	0,712	0,709	0,711	0,711
led7digit	0,727	0,729	0,690	0,726	0,722	0,710
musk	0,760	0,762	0,778	0,768	0,764	0,773
new-thyroid1	0,943	0,943	0,956	0,948	0,946	0,959
pima	0,771	0,770	0,766	0,770	0,773	0,766
sonar	0,780	0,778	0,794	0,776	0,781	0,788
transfusion	0,780	0,782	0,786	0,787	0,783	0,789
vehicle	0,752	0,752	0,757	0,753	0,751	0,756
vehicle2	0,946	0,945	0,949	0,946	0,946	0,948
vowel	0,963	0,962	0,973	0,963	0,965	0,972
vowel0	0,963	0,963	0,973	0,963	0,965	0,972
wdbc	0,967	0,967	0,967	0,967	0,967	0,967
wine	0,971	0,970	0,973	0,970	0,971	0,970
yeast-2_vs_4	0,950	0,950	0,952	0,950	0,951	0,951
ranking	4,32	4,33	2,53	3,35	3,80	2,67
p-value	0,002	0,001	0,573	0,068	0,002	-

Tabela 29 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (LR). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento médio para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,615	0,602	0,699	0,696	0,644	0,726
banana	0,572	0,569	0,632	0,639	0,581	0,635
blood	0,539	0,537	0,560	0,580	0,548	0,585
bupa	0,632	0,628	0,641	0,647	0,639	0,642
cleveland	0,337	0,330	0,352	0,335	0,332	0,341
cmc	0,467	0,466	0,477	0,473	0,471	0,476
column_3C	0,804	0,805	0,801	0,803	0,802	0,801
dermatology	0,965	0,965	0,965	0,963	0,965	0,966
diabetes	0,729	0,729	0,728	0,729	0,729	0,728
ecoli-0_vs_1	0,958	0,958	0,960	0,958	0,956	0,960
german	0,630	0,629	0,640	0,635	0,629	0,640
glass1	0,411	0,421	0,579	0,582	0,501	0,552
glass6	0,855	0,847	0,886	0,880	0,877	0,888
haberman	0,494	0,491	0,525	0,537	0,505	0,557
hayes	0,604	0,580	0,676	0,625	0,626	0,677
heart	0,821	0,822	0,824	0,823	0,822	0,823
ilpd	0,486	0,486	0,548	0,533	0,515	0,547
led7digit	0,720	0,721	0,673	0,719	0,714	0,701
musk	0,750	0,752	0,771	0,759	0,755	0,764
new-thyroid1	0,873	0,873	0,910	0,888	0,879	0,916
pima	0,729	0,729	0,725	0,730	0,732	0,726
sonar	0,777	0,774	0,790	0,772	0,777	0,785
transfusion	0,535	0,537	0,567	0,582	0,553	0,590
vehicle	0,744	0,745	0,750	0,745	0,744	0,749
vehicle2	0,929	0,928	0,933	0,929	0,929	0,931
vowel	0,873	0,873	0,913	0,876	0,881	0,909
vowel0	0,874	0,874	0,913	0,875	0,883	0,909
wdbc	0,964	0,964	0,963	0,964	0,963	0,963
wine	0,970	0,970	0,972	0,970	0,971	0,970
yeast-2_vs_4	0,824	0,824	0,838	0,825	0,831	0,831
ranking	4,48	4,87	2,32	3,05	3,90	2,38
p-value	0,000	0,000	0,444	0,003	0,000	-

Tabela 30 – Avaliação do MCC comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (LR). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-value apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,325	0,307	0,483	0,477	0,393	0,543
banana	0,344	0,346	0,423	0,425	0,357	0,421
blood	0,215	0,215	0,242	0,267	0,226	0,271
bupa	0,318	0,311	0,305	0,329	0,322	0,309
cleveland	0,360	0,346	0,345	0,352	0,344	0,349
cmc	0,216	0,216	0,228	0,223	0,222	0,224
column_3C	0,760	0,761	0,757	0,759	0,758	0,758
dermatology	0,962	0,962	0,962	0,959	0,962	0,963
diabetes	0,473	0,471	0,471	0,470	0,472	0,468
ecoli-0_vs_1	0,920	0,921	0,924	0,920	0,917	0,924
german	0,318	0,319	0,332	0,326	0,316	0,334
glass1	0,000	0,030	0,236	0,270	0,156	0,220
glass6	0,735	0,719	0,784	0,777	0,773	0,787
haberman	0,110	0,102	0,155	0,176	0,118	0,198
hayes	0,378	0,340	0,507	0,409	0,412	0,501
heart	0,651	0,654	0,657	0,654	0,654	0,656
ilpd	0,098	0,109	0,169	0,159	0,143	0,175
led7digit	0,698	0,700	0,660	0,697	0,693	0,681
musk	0,507	0,510	0,546	0,524	0,516	0,535
new-thyroid1	0,776	0,778	0,840	0,802	0,787	0,849
pima	0,473	0,472	0,462	0,473	0,479	0,465
sonar	0,563	0,559	0,588	0,553	0,564	0,580
transfusion	0,207	0,216	0,249	0,266	0,231	0,276
vehicle	0,674	0,674	0,680	0,675	0,673	0,678
vehicle2	0,859	0,857	0,868	0,859	0,860	0,864
vowel	0,762	0,761	0,835	0,766	0,777	0,828
vowel0	0,763	0,764	0,837	0,765	0,779	0,828
wdbc	0,928	0,928	0,927	0,928	0,926	0,926
wine	0,956	0,955	0,960	0,955	0,958	0,955
yeast-2_vs_4	0,686	0,688	0,707	0,687	0,699	0,695
ranking	4,45	4,50	2,48	3,22	3,80	2,55
p-value	0,000	0,000	0,444	0,008	0,000	-

Tabela 31 – Avaliação da taxa de acerto média comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,796	0,797	0,832	0,828	0,811	0,835
banana	0,654	0,653	0,695	0,689	0,656	0,686
blood	0,775	0,775	0,778	0,785	0,777	0,784
bupa	0,631	0,626	0,654	0,663	0,643	0,670
cleveland	0,609	0,604	0,598	0,605	0,604	0,608
cmc	0,456	0,457	0,488	0,484	0,466	0,477
column_3C	0,848	0,848	0,843	0,846	0,848	0,850
dermatology	0,969	0,970	0,967	0,970	0,968	0,968
diabetes	0,751	0,749	0,748	0,758	0,754	0,756
ecoli-0_vs_1	0,967	0,966	0,971	0,967	0,967	0,972
german	0,709	0,706	0,720	0,723	0,717	0,723
glass1	0,646	0,649	0,684	0,691	0,66	0,685
glass6	0,906	0,902	0,938	0,918	0,913	0,937
haberman	0,725	0,724	0,725	0,733	0,727	0,732
hayes	0,585	0,563	0,676	0,609	0,618	0,663
heart	0,829	0,828	0,823	0,831	0,823	0,831
ilpd	0,712	0,711	0,710	0,707	0,71	0,711
led7digit	0,726	0,728	0,691	0,725	0,717	0,714
musk	0,764	0,762	0,783	0,773	0,767	0,777
new-thyroid1	0,942	0,940	0,951	0,944	0,945	0,953
pima	0,751	0,751	0,752	0,758	0,751	0,755
sonar	0,775	0,774	0,788	0,774	0,777	0,780
transfusion	0,774	0,774	0,777	0,782	0,776	0,781
vehicle	0,763	0,762	0,767	0,762	0,76	0,767
vehicle2	0,950	0,950	0,955	0,951	0,95	0,953
vowel	0,952	0,950	0,975	0,954	0,958	0,972
vowel0	0,952	0,951	0,975	0,953	0,959	0,970
wdbc	0,966	0,966	0,966	0,967	0,965	0,966
wine	0,972	0,971	0,972	0,971	0,973	0,971
yeast-2_vs_4	0,937	0,936	0,944	0,942	0,939	0,945
ranking	4,30	4,93	2,93	2,90	3,82	2,12
p-value	0,000	0,000	0,107	0,020	0,000	-

Tabela 32 – Avaliação do F-score médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento médio para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

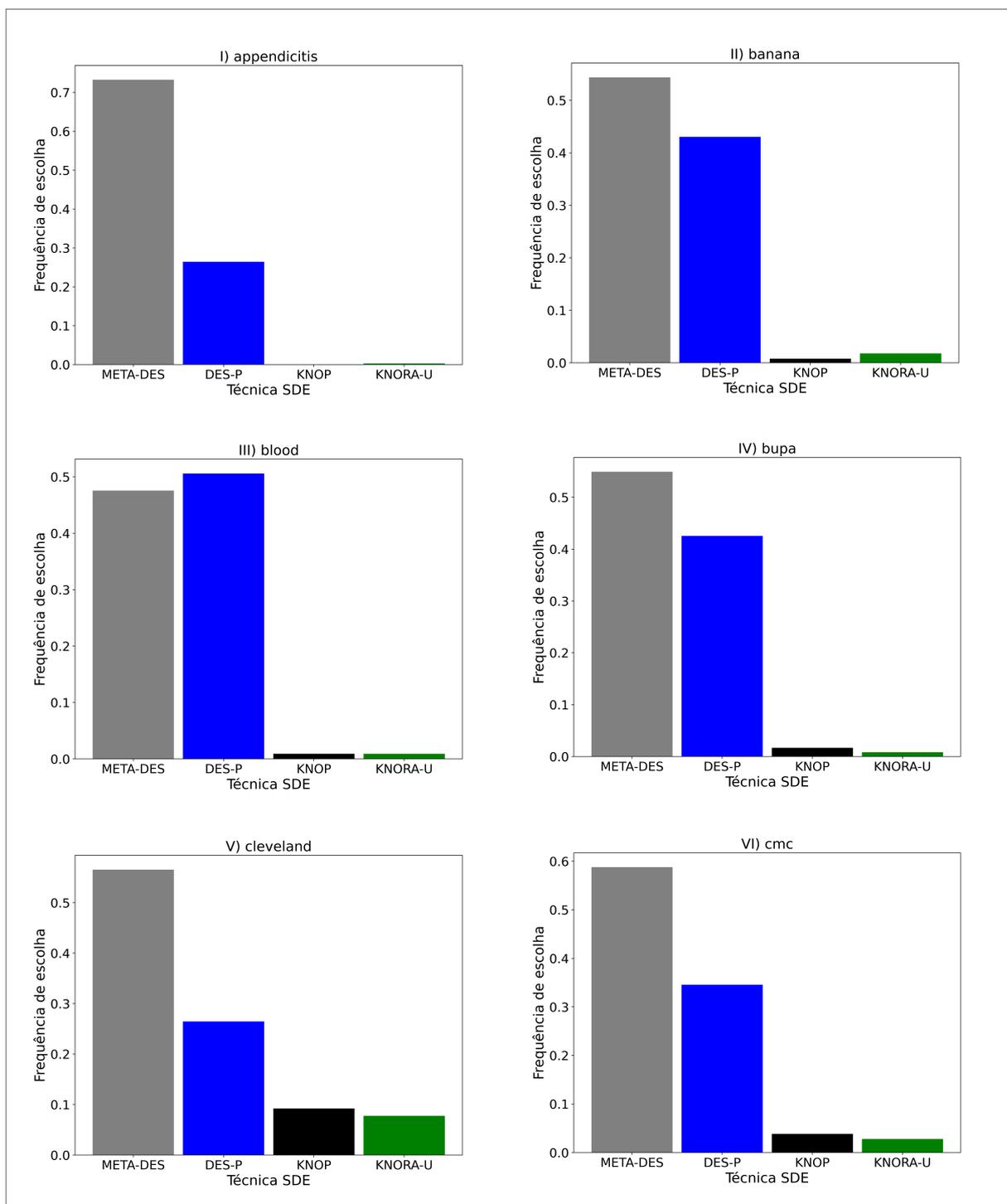
Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,480	0,483	0,628	0,621	0,558	0,647
banana	0,570	0,569	0,640	0,634	0,576	0,630
blood	0,477	0,478	0,516	0,545	0,497	0,549
bupa	0,518	0,496	0,622	0,606	0,573	0,625
cleveland	0,319	0,312	0,327	0,314	0,321	0,331
cmc	0,312	0,309	0,413	0,391	0,354	0,391
column_3C	0,795	0,794	0,789	0,794	0,797	0,801
dermatology	0,965	0,966	0,964	0,966	0,965	0,964
diabetes	0,672	0,666	0,683	0,696	0,682	0,697
ecoli-0_vs_1	0,963	0,962	0,967	0,963	0,963	0,969
german	0,460	0,449	0,534	0,529	0,501	0,532
glass1	0,400	0,410	0,587	0,576	0,497	0,562
glass6	0,743	0,718	0,851	0,787	0,762	0,845
haberman	0,432	0,429	0,488	0,508	0,463	0,522
hayes	0,602	0,577	0,679	0,629	0,629	0,670
heart	0,824	0,822	0,818	0,826	0,817	0,826
ilpd	0,417	0,415	0,522	0,506	0,469	0,510
led7digit	0,721	0,723	0,676	0,719	0,711	0,705
musk	0,753	0,750	0,775	0,763	0,757	0,768
new-thyroid1	0,872	0,865	0,896	0,879	0,879	0,900
pima	0,672	0,670	0,690	0,696	0,678	0,696
sonar	0,771	0,769	0,784	0,770	0,772	0,776
transfusion	0,475	0,473	0,511	0,536	0,498	0,541
vehicle	0,758	0,757	0,764	0,757	0,755	0,764
vehicle2	0,933	0,933	0,940	0,935	0,934	0,938
vowel	0,820	0,808	0,916	0,828	0,848	0,904
vowel0	0,820	0,813	0,916	0,827	0,852	0,899
wdbc	0,962	0,962	0,962	0,964	0,961	0,962
wine	0,972	0,970	0,972	0,970	0,973	0,971
yeast-2_vs_4	0,757	0,754	0,795	0,782	0,766	0,799
ranking	4,57	5,23	2,33	3,13	3,83	1,90
p-value	0,000	0,000	0,329	0,000	0,000	-

Tabela 33 – Avaliação do MCC médio comparando a abordagem PS-DES com quatro abordagens SDE (CRUZ et al., 2015): KNOP, KNORA-U, META-DES e DES-P e o PS-Aleatório. Todas as abordagens utilizam um *pool* homogêneo com 300 classificadores-base (Perceptron). O melhor resultado de cada conjunto de dados é apresentado em negrito. A linha *ranking* apresenta o ranqueamento média para cada técnica. A linha p-valor apresenta o resultado da aplicação do teste estatístico Wilcoxon pareado quando comparado PS-DES com a técnica apresentada na coluna.

Base	KNOP	KNORA-U	META-DES	DES-P	PS-Aleatório	PS-DES
appendicitis	0,069	0,076	0,361	0,341	0,229	0,395
banana	0,373	0,373	0,440	0,422	0,37	0,419
blood	0,135	0,135	0,182	0,231	0,166	0,231
bupa	0,220	0,201	0,272	0,295	0,242	0,311
cleveland	0,321	0,309	0,302	0,311	0,31	0,322
cmc	0,109	0,111	0,180	0,169	0,131	0,157
column_3C	0,763	0,762	0,753	0,758	0,761	0,764
dermatology	0,961	0,963	0,959	0,962	0,96	0,960
diabetes	0,419	0,414	0,410	0,437	0,428	0,432
ecoli-0_vs_1	0,928	0,927	0,936	0,928	0,929	0,939
german	0,140	0,115	0,213	0,229	0,197	0,226
glass1	0,017	0,047	0,251	0,274	0,155	0,247
glass6	0,546	0,505	0,722	0,621	0,582	0,718
haberman	0,009	0,001	0,101	0,135	0,072	0,147
hayes	0,371	0,336	0,511	0,415	0,422	0,495
heart	0,659	0,657	0,649	0,665	0,646	0,665
ilpd	0,007	0,000	0,139	0,129	0,092	0,130
led7digit	0,696	0,699	0,660	0,696	0,686	0,684
musk	0,515	0,511	0,555	0,534	0,523	0,542
new-thyroid1	0,776	0,764	0,815	0,787	0,787	0,824
pima	0,419	0,420	0,423	0,437	0,419	0,430
sonar	0,554	0,549	0,577	0,549	0,554	0,563
transfusion	0,130	0,128	0,172	0,213	0,162	0,218
vehicle	0,687	0,686	0,692	0,686	0,683	0,692
vehicle2	0,869	0,869	0,882	0,873	0,871	0,877
vowel	0,685	0,664	0,846	0,697	0,729	0,826
vowel0	0,683	0,672	0,845	0,694	0,736	0,816
wdbc	0,925	0,925	0,924	0,928	0,922	0,925
wine	0,959	0,956	0,959	0,957	0,96	0,958
yeast-2_vs_4	0,591	0,588	0,645	0,628	0,603	0,651
ranking	4,33	5,00	2,77	2,87	4,03	2,00
p-value	0,000	0,000	0,111	0,007	0,000	-

APÊNDICE B – ESCOLHAS DAS TÉCNICAS DE SELEÇÃO DINÂMICA DE CLASSIFICADORES PELO PS-DES PARA TODAS AS BASES AVALIADAS

Figura 9 – Frequência percentual das escolhas das técnicas SDE realizadas pelo PS-DES na Etapa 2.4., para as 30 bases de dados



Fonte: Elaborada pelo autor(2024)

