

# Meta aprendizado aplicado ao câncer de mama

Pedro Henrique Tonidandel<sup>1</sup>, Ronan Assumpção Silva<sup>1,2</sup>, André Luiz Brun<sup>3</sup>

<sup>1</sup> Instituto Federal do Paraná (IFPR) – Pinhais – PR – Brasil

<sup>2</sup> Universidade Federal do Paraná (UFPR) – Curitiba – PR – Brasil

<sup>3</sup> Universidade Estadual do Oeste do Paraná (UNIOESTE) – Cascavel – PR – Brasil

pedrohdandel@gmail.com, ronan.silva@ifpr.edu.br, andre.brun@unioeste.br

**Abstract.** *Among the techniques to assist in the disease detection process, we can mention the Classification, which consists of labeling an instance based on the already known categories of the investigated problem. In this proposal, we consider the detection of breast cancer as a classification problem. We use Multiple Classifier Systems, to generate several classifiers, functioning as computational specialists - analogous to the consultation with several specialists in the health area. Based on the opinion of the specialists, a metabase of the problem was created, which, when treated as a new classification problem, allowed the evaluation of patterns in the opinions that affect the correctness of the labeling. The proposal is, on average,  $\approx 15$  percentage points higher than the literature.*

**Resumo.** *Dentre as técnicas para auxiliar no processo de detecção de doenças, podemos citar a Classificação, que consiste em rotular uma instância baseada nas categorias já conhecidas do problema investigado. Nesta proposta, consideramos a detecção do câncer de mama como um problema de Classificação. Usamos Sistemas de Múltiplos Classificadores, para gerar diversos classificadores, funcionando como especialistas computacionais - análoga a consulta com diversos especialistas da área da saúde. A partir da opinião dos especialistas foi criada uma metabase do problema, que, ao ser tratada como um novo problema de classificação, permitiu avaliar padrões nas opiniões que afetam o acerto na rotulação. A proposta tem aumentado o acerto de outras técnicas da literatura - considerando a média - em  $\approx 15$  pontos percentuais.*

## 1. Introdução

O Ministério da Saúde aponta que há um aumento considerável no número de exames para detecção do Câncer de Mama nos últimos 13 anos [Ministério da Saúde 2022]. Essa é uma notícia muito importante, pois o diagnóstico precoce é essencial para aumentar as taxas de sucesso na escolha do tratamento que, afeta também na qualidade de vida do paciente. Dentre os métodos mais utilizados está o rastreamento, que analisa as características da mama e tenta detectar a presença do tumor antes da ocorrência de sintomas. Um dos tipos de rastreamento mais utilizados, segundo a Organização Mundial da Saúde, é a mamografia. Entretanto, é comum exames das mais variadas técnicas e objetivos estarem suscetíveis à presença de inconsistências, tal como a ocorrência de falsos positivos ou falsos negativos. Na intenção de evitar diagnósticos incorretos, faz-se necessária uma maneira de analisar os dados provenientes dos exames. Isso beneficia as chances de um diagnóstico

correto, além de auxiliar na escolha dos tratamentos. Diante desse cenário, é desejável uma ferramenta que, disponível ao profissional da área médica, auxilie nas análises.

Como área, o Reconhecimento de Padrões (RP) se apresenta como uma das alternativas mais robustas para a análise dos dados. Dentre as aplicações que empregam técnicas de RP estão a detecção de doenças, detecção de fraudes em cartão de crédito, pontuação em crédito, classificação de sequências de DNA, reconhecimento facial, entre outras [Bashbaghi. et al. 2017, Cruz et al. 2018]. O funcionamento desses sistemas requerem amostras, que atuam como uma fonte de entrada de dados. Na Classificação supervisionada há a fase de treino e de teste. Na primeira fase, o objetivo é encontrar os padrões entre esses dados e a categoria, que é conhecida. Na próxima fase, também conhecida como previsão, as amostras ainda não tem uma categoria associada, portanto, cada amostra é rotulada a partir dos padrões encontrados na fase de treinamento.

Para evitar que um classificador tenha que descobrir todos os padrões possíveis, o que é geralmente inviável, surge como alternativa os Sistemas de Múltiplos Classificadores (SMC) [Silva et al. 2020]. Nesses sistemas, um grupo de classificadores é treinado sob diferentes amostras (grupo homogêneo) ou um grupo de distintos classificadores é treinado com as mesmas amostras (grupo heterogêneo). Esses sistemas são semelhantes a coleta de opinião perante vários especialistas antes de tomar uma decisão. Em um dado cenário, a decisão pode estar inclinada a seguir com a maioria dos especialistas (classificadores) ou ela deve ser baseada em um contexto, onde alguns especialistas são tecnicamente mais competentes [Brun et al. 2018].

Um SMC é composto por fases distintas, que inicia com a criação dos classificadores (especialistas computacionais), inclui ou não uma fase de seleção dos classificadores e, por fim, há a combinação de votos entre os especialistas, emitindo um parecer do sistema. Porém, há de ser levado em consideração não somente a resposta final do sistema, mas também as limitações desses especialistas. Neste artigo, o voto entre os classificadores é tido como um novo problema de classificação. Com base no voto dos classificadores em uma base de validação, é construída uma metabase de dados. A partir dela é possível investigar a habilidade dos classificadores.

Este estudo está organizado da seguinte forma: Na Seção 2, apresenta-se o referencial teórico, que traz apenas os conceitos principais para o entendimento da proposta. A metodologia, na qual detalhamos a proposta de classificação é apresentada na Seção 3. Por fim, os experimentos, resultados e conclusões são detalhados nas Seções 4, 5 e 6.

## **2. Referencial Teórico**

Esta seção apresenta dados importantes sobre o câncer de mama no Brasil, as diferenças entre as regiões e o que tem sido feito recentemente para melhorar o cenário. Em seguida, descreve-se as técnicas de RP fundamentais para o entendimento da nossa proposta.

### **2.1. Sobre o câncer de mama**

Este tipo de câncer representou, em 2020, 24,5% dos casos de câncer em mulheres e, estimou-se que ele foi o causador de cerca de 685 mil mortes ao redor do mundo [IARC 2020]. Entretanto, este número vem diminuindo. Em [Tabár et al. 2019] é apontado que a diminuição do risco de mortalidade entre mulheres que fazem a mamografia

com frequência em comparação às que não fazem é de 60%, no período mais recente da pesquisa que envolveu 10 anos de estudo.

A distribuição de profissionais com expertise na identificação do câncer no Brasil é conhecidamente desigual [Mattos et al. 2013]. Estudos realizados no ano de 2008, na região sul do Brasil, indicam que cerca de 15,9% das mulheres nunca realizou o exame clínico das mamas ou da mamografia. Na região nordeste esta taxa é maior, alcançando 29,7% [da Silva Borges et al. 2016]. Dentre os problemas apontados está a falta de mamógrafos, a distância até o local do exame, a oferta de tratamentos e a escassez de especialistas.

Portanto, se faz necessária uma estratégia para que o diagnóstico de doenças ocorra o mais rápido possível. Um dos recursos que tem sido empregados nos últimos anos é o da Telemedicina. Esta só foi possível pela crescente oferta de infraestrutura para Internet, especialmente em localidades mais afastadas. Equipamentos disponibilizados em laboratório passaram a ser levados até as pessoas por embarcações simples, graças ao tamanho reduzido que a nanotecnologia tem proporcionado. Além da disponibilidade de equipamentos e de consultas remotas, é importante mencionar que técnicas de RP auxiliam os profissionais da saúde a algum tempo [Murphy 2012, Fusco et al. 2016].

## 2.2. Classificação

A Classificação consiste na tentativa de aprender o relacionamento entre conjuntos de características e rótulos previamente definidos e, a partir dessa perspectiva, tentar prever corretamente a classe de objetos apresentados ao sistema que ainda não possuem rótulo [Aggarwal 2015]. O algoritmo de classificação Perceptron, proposto por [Rosenblatt 1958], é um dos mais simples modelos de redes neurais artificiais, o qual simula o funcionamento de um neurônio humano. Consiste basicamente em nós de entrada e de saída, onde o primeiro representa os atributos de entrada e o segundo a saída do modelo. A cada atributo  $i$  é atribuído um peso  $w_i$  que é multiplicado pelo valor de entrada  $x_i$  do atributo. Os resultados de todas as ponderações são somadas a um bias, gerando uma saída  $y$  que corresponde a um valor binário (neurônio ativo ou inativo). Essas características fazem com que o Perceptron seja um classificador binário, conseguindo distinguir um problema limitado a apenas duas classes. Na implementação de [Freund and Shapire 1999], o VotedPerceptron é considerado uma versão mais rápida do algoritmo Perceptron clássico de [Rosenblatt 1958] e recomendado para problemas cujas classes são separáveis linearmente.

## 2.3. Sistema de Múltiplos Classificadores

Muitas vezes os problemas de classificação apresentam grandes variações entre os objetos. Sendo assim, um classificador sozinho (sistema monolítico) tem dificuldades de interpretar essas variações satisfatoriamente. Uma solução para enfrentar tal desafio é a adoção de diversos classificadores com o objetivo de tornar o sistema mais robusto [Brun et al. 2018]. Assim, quando há diversidade, os erros de um classificador são supridos por um ou mais membros do conjunto, para então, como equipe, aumentarem a taxa de reconhecimento. Quando comparados a sistemas monolíticos, esses sistemas dificilmente apresentam resultados inferiores [Brun et al. 2018, Silva et al. 2020].

Um SMC pode ser dividido nas fases de Geração, Seleção e Combinação [Silva 2018]. Na primeira, cria-se um grupo de classificadores precisos e diferentes entre

si, como faz o Bagging (*Bootstrap Aggregating*) [Breiman 1996]. Na segunda, é reduzido o conjunto inicial de classificadores com o objetivo de manter ou aumentar a acurácia do sistema. Já na fase de combinação, regras para considerar os votos dos classificadores são adotadas para que o sistema tenha a sua resposta [Nagi and Bhattacharyya 2013].

## 2.4. Metalearning

Enquanto a Aprendizagem de Máquina (AM) tradicional trabalha sobre um conjunto de dados por vez, a meta aprendizagem é baseada no acúmulo de experiência do desempenho de múltiplas aplicações de um algoritmo de AM [Souza 2010]. O conhecimento no Meta-Aprendizado é obtido a partir de um conjunto de meta-exemplos. Nele é armazenada a experiência obtida na aplicação de um conjunto de algoritmos candidatos em problemas de aprendizado investigados em momentos anteriores. O conjunto da experiência acumulada é chamado metabase e os elementos são chamados de meta exemplos. Estes, são empregados pelo meta aprendiz para relacionar características do problema com o desempenho do algoritmo candidato [Prudêncio and Ludermir 2007].

O aprendizado em *meta level* preocupa-se tanto com o acúmulo de experiências quanto ao próprio processo de aprendizagem [Ferrari 2014]. A composição de um meta-classificador pode ser considerada a partir de 4 passos [Liu et al. 2018]. Inicialmente os algoritmos de aprendizado são treinados em um conjunto de dados de treino. Na etapa seguinte um conjunto de validação é usado para avaliar os classificadores treinados anteriormente. No terceiro passo então, as classes que foram atribuídas corretamente são usadas para treinar um meta-algoritmo. No último passo um meta-classificador é produzido e o processo de meta aprendizado é concluído.

Uma das abordagens mais conhecidas que aplica o *metalearning* é o Stacking. Este realiza o mapeamento dos metadados para a previsão de classes ou de probabilidades, explorando as diferenças entre os classificadores [Brazdil et al. 2009] e construindo modelos preditivos sobre a classificação de algoritmos base [Whalen and Pandey 2013].

## 3. Metodologia

Nesta seção é apresentada uma nova proposta para a classificação de dados. Este método envolve o uso de uma metabase de dados sobre um problema que permita a técnica de Classificação. Neste estudo considera-se bases de dados cujas classes recaem na presença ou ausência de nódulo maligno, referente ao câncer de mama.

### 3.1. Bases de dados

As três bases com dados referentes ao diagnóstico do câncer de mama estão disponíveis no repositório UCI Machine Learning [Dua and Graff 2019]. São elas: i) Breast Cancer Coimbra, ii) Mammographic Mass e iii) Wisconsin Breast Cancer (Diagnose).

Uma forma de expressar a distribuição das instâncias dentro das classes presentes em cada conjunto de dados é o *Imbalance Ratio* (IR) ou Taxa de Desequilíbrio. O índice é obtido pela divisão da quantidade de instâncias da classe mais frequente pela menos frequente na base de dados. Valores próximos de 1 indicam que a proporcionalidade é respeitada. Por outro lado, valores maiores indicam que há classes com maior presença no conjunto de dados do que outras [Silva et al. 2020].

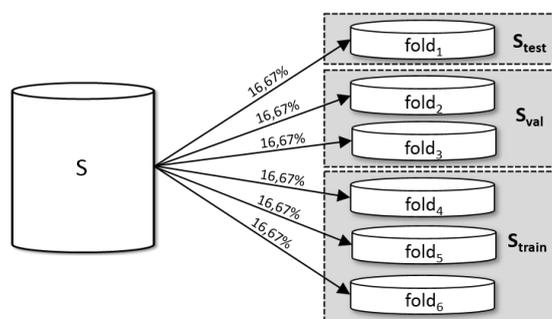
Uma síntese das características das bases utilizadas neste trabalho são apresentadas na Tabela 1. Base corresponde a abreviação, sendo Breast Cancer Coimbra (BCC), Mammographic Mass (MM) e Wisconsin Breast Cancer (Diagnose) (WDBC). Instâncias correspondem ao total de amostras. Treino, Teste e Validação correspondem à distribuição do total de instâncias para cada fase. Atributos refere-se aos valores observados e classes ao número de categorias. A coluna “R” indica a taxa de desequilíbrio (*imbalance ratio*). Todas essas bases de dados são encontradas na UCI [Dua and Graff 2019].

**Tabela 1. Principais características das bases usadas nos experimentos**

Base	Instâncias	Treino	Teste	Validação	Atributos	Classes	IR
BCC	116	58	19	39	10	2	1,23
MM	830	413	139	278	6	2	1,06
WDBC	569	284	95	190	31	2	1,68

### 3.2. Classificação por MetaDados - MD

Os conjuntos de dados foram divididos em 6 folds exatamente proporcionais e estratificados conforme ilustrado na Figura 1. A base Mammographic Mass, por exemplo, tem em cada porção, 71 elementos referentes à classe benigna e 68 à classe maligna. Para a etapa de treinamento dos classificadores, 3 folds (50% da base original) foram utilizados (identificado como  $S_{train}$ ). Já para o estágio da validação e construção da metabase adotou-se 2 dos seis folds (33,33% das instâncias da base, formando o conjunto  $S_{val}$ ) e, para o teste, foi empregado o fold restante ( $S_{test}$ ), que corresponde a aproximadamente 16,67% do total de registros da base original.

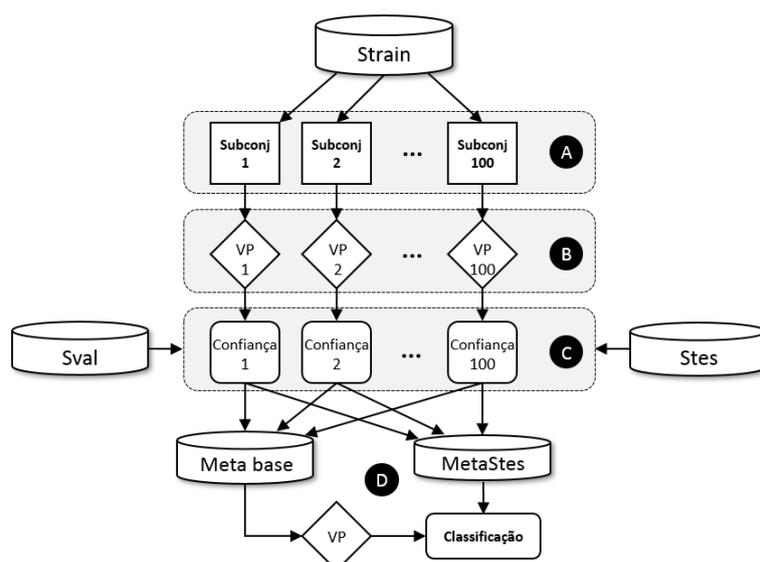


**Figura 1. Divisão da base original em treino, teste e validação**

O Bagging foi escolhido para construir um conjunto de 100 subconjuntos aleatoriamente a partir do conjunto  $S_{train}$ . Cada subconjunto gerado contém aproximadamente 66% do tamanho de  $S_{train}$ , ou seja, cerca de 33% do tamanho do conjunto original  $S$ .

Cada subconjunto formado pelo Bagging foi usado no treinamento de um classificador distinto. Dessa forma, cada um dos 100 conjuntos homogêneos foi utilizado para a aprendizagem de um Voted Perceptron (VP), proposto por [Freund and Shapire 1999]. Assim, com essa estratégia, espera-se que seja estimulada a diversidade de opiniões entre os classificadores treinados nesses subconjuntos diversificados.

A Figura 2 ilustra a etapa de geração da metabase. Inicialmente (Figura 2-A) são gerados os 100 subconjuntos por Bagging. Em seguida, cada subconjunto é usado no treinamento de um classificador (Figura 2-B) do tipo VP).



**Figura 2. Geração de classificadores para o algoritmo proposto**

A opinião de cada classificador sobre todas as instâncias do conjunto de validação ( $S_{val}$ ) é coletada como ilustrado na Figura (2-C). Ao dar a sua opinião, cada classificador define um grau de confiança (que varia entre 0 e 1) no seu voto. Portanto, para cada instância do conjunto de validação tem-se um total de 100 índices de confiança. Esses valores são então utilizados como metacaracterísticas e, junto com a classe original da instância, formam um vetor composto de 101 atributos, utilizado para representar apenas um registro na metabase, ou seja, um registro de metainformações sobre uma instância de validação particular. Conforme representado na Figura (2-C), o conjunto de  $S_{test}$  é submetido ao mesmo processo de extração de metacaracterísticas. Assim, para a etapa de avaliação do modelo é adotado o conjunto Meta  $S_{test}$ , que é composto pelos mesmo 101 atributos da metabase utilizada para treino.

Na etapa seguinte os registros da metabase são usados para treinar um classificador monolítico (Figura 2-D). O método de aprendizagem empregado nesta etapa é o mesmo do treinamento dos classificadores do *pool* inicial (VP). Este classificador é então responsável pela rotulação das instâncias pertencentes ao conjunto  $S_{test}$ .

## 4. Experimentos

Visando estimar a precisão do método proposto adotou-se a estratégia de validação cruzada. Uma vez que o conjunto de dados foi dividido em 6 folds, o processo de avaliação foi realizado seis vezes. Como esse processo separa um fold para teste e 5 para o treinamento e validação, foram separadas as instâncias aleatoriamente para compor os três folds referentes a  $S_{train}$  e os dois restantes que formam  $S_{val}$ . A técnica proposta, chamada de MD faz uso de um algoritmo de classificação para o treinamento dos subconjuntos de dados e do aprendizado da metabase. Desta forma, é interessante a comparação do resultado com um classificador monolítico. O algoritmo de classificação usado, VP, teve todos os parâmetros configurados como o default da ferramenta Weka<sup>1</sup> na versão 3.9.2.

<sup>1</sup><https://www.cs.waikato.ac.nz/ml/weka/>

A Tabela 2 demonstra os resultados obtidos em cada fold com relação à taxa de acerto, com média aritmética (MA) e desvio padrão (DP). Percebe-se que a técnica proposta neste trabalho é, em média, superior ao classificador monolítico para as três bases testadas. Em negrito estão os melhores resultados da comparação.

**Tabela 2. Resultados obtidos com classificadores monolíticos (VP) e com base na estratégia proposta ( $MD^{VP}$ ).**

Fold \ Dataset	BCC		MM		WDBC	
	VP	$MD^{VP}$	VP	$MD^{VP}$	VP	$MD^{VP}$
0	55,00	<b>60,00</b>	48,92	<b>72,66</b>	<b>92,63</b>	91,58
1	<b>60,00</b>	50,00	49,64	<b>76,98</b>	<b>86,32</b>	82,42
2	47,37	<b>57,89</b>	72,46	<b>76,09</b>	69,47	<b>94,74</b>
3	52,63	<b>63,16</b>	66,67	<b>75,36</b>	<b>87,37</b>	83,16
4	<b>57,89</b>	42,11	63,77	<b>77,54</b>	87,37	<b>91,58</b>
5	57,89	<b>58,97</b>	64,49	<b>74,64</b>	<b>88,30</b>	87,23
MA	55,13	<b>55,36</b>	60,99	<b>75,55</b>	85,24	<b>88,45</b>
DP	04,19	07,14	08,74	01,61	07,34	04,56

Para analisar a diferença entre as técnicas, a matriz de confusão pode ser um parâmetro interessante. Esta estratégia representa as classes reais e previstas em uma matriz quadrada onde cada linha corresponde a uma classe esperada, enquanto cada coluna corresponde à classe prevista. Quando o classificador acerta uma instância, então ela é representada na linha e coluna corretas, que correspondem à diagonal principal. Dentre as vantagens em adotar a matriz de confusão há a possibilidade de verificar a tendência dos classificadores monolíticos em decidir mais vezes por uma das classes, por exemplo. Portanto, para as devidas análises são apresentadas as respectivas matrizes nas Tabelas 3a, 3b e 3c.

**Tabela 3. Tendências dos comportamentos dos classificadores ilustrados através de matrizes de confusão**

Real \ Previsto	VP		$MD^{VP}$	
	C1	C2	C1	C2
c1	1	7	2	6
c2	1	10	5	6

(a) Matriz de confusão correspondente a base BCC no fold 4.

Real \ Previsto	VP		$MD^{VP}$	
	C1	C2	C1	C2
c1	2	70	49	23
c2	0	67	9	58

(b) Matriz de confusão correspondente a base MM no fold 1.

Real \ Previsto	VP		$MD^{VP}$	
	C1	C2	C1	C2
c1	31	29	59	1
c2	0	35	4	31

(c) Matriz de confusão correspondente a base WDBC no fold 2.

Na Tabela 3a percebe-se a tendência de VP classificar os padrões como sendo prioritariamente da classe 2. Como a maior parte das instâncias pertencem a essa classe, o VP acerta devido a essa tendência. Por outro lado,  $MD^{VP}$ , apesar de menos tendencioso, tem menos acertos como um todo.

A tendência de VP em classificar os padrões como sendo quase sempre da classe 2 também é observada na Tabela 3b. Por outro lado,  $MD^{VP}$  é capaz de sugerir a classe

dos padrões de maneira mais apropriada.

A aplicação das duas técnicas na base WDBC estão na Tabela 3c. A distribuição das opiniões do monolítico continua apresentando tendência maior para uma das classes, embora seja menor nesta base. Por outro lado, as opiniões de  $MD^{VP}$  apresentam uma distribuição bastante próxima do ideal, pois das 95 instâncias, errou apenas 5.

## 5. Resultados e Discussão

Considerando apenas a média, o monolítico, identificado como  $VP$ , perdeu para  $MD^{VP}$  em todas as bases. A diferença média mais significativa foi observada na base MM, com mais de 14 pontos percentuais de diferença. Além disso, o monolítico perdeu em acurácia na maioria das comparações entre os folds.

No caso da base BCC percebe-se, a partir da matriz de confusão, a tendência de  $VP$  em considerar as observações como sendo apenas pertencentes a classe 2. Considerando o panorama geral das duas técnicas, observa-se que a falta de observações e a separabilidade de classes foram fatores que levaram ao insucesso, seja por tendência ou pela falta dela.

A base MM contém a maior diferença média entre os métodos, considerando a vantagem de aproximadamente 15 pontos percentuais do  $MD^{VP}$  (Tabela 2) frente ao  $VP$ . A partir da Tabela 3b, percebe-se a distribuição das previsões de maneira mais balanceada em relação as outras bases e também em relação ao monolítico.

Em relação a base WDBC, a nova abordagem possui uma capacidade de reconhecimento atrativa, com destaque a diferença observada no fold 2. Enquanto o monolítico continua a apresentar tendência por uma das classes,  $MD$  distribuiu as opiniões corretamente na maioria das observações submetidas ao sistema classificador.

Diante das três bases estudadas, é importante lembrar do algoritmo de classificação utilizado. Ao considerar que  $VP$  é um classificador para classes separáveis linearmente, é incapaz de aprender os padrões apresentados pela base BCC para ambas as técnicas. Além disso, a quantidade baixa de observações não promove um aprendizado adequado para a construção de um modelo robusto. Nas bases MM e WDBC, há uma considerável quantidade de observações disponíveis para o treinamento da abordagem monolítica, bem como o treino e a validação da proposta chamada  $MD$ . Essa quantidade de dados aumentou consideravelmente as chances de sucesso das abordagens, sendo que, em média,  $MD$  é um classificador mais acurado e com menor desvio padrão, o que sugere ser um modelo mais confiável que um classificador monolítico.

## 6. Conclusão

Esta pesquisa propôs o uso de uma metabase para a melhoria de um sistema classificador. Para isso, mesclou conceitos de Sistemas de Múltiplos Classificadores e Metalearning. A ideia é inspirada na consulta a múltiplos especialistas, considerando tanto sua opinião quanto a confiança em que eles têm de estarem certos. Aqui, a confiança dos classificadores dentro de um SMC é tida como um novo problema de classificação. Isto permite ir além da análise do padrão, considerando que a confiança e expertise dos especialistas sobre um dado problema também deve influenciar a decisão final do sistema.

Com base na confiança dos classificadores, uma metabase de opiniões foi criada. Ao usar a metabase como um novo problema de classificação, tem-se como ponto forte

explorar as vantagens dos classificadores enquanto evita as suas fraquezas. Esse processo está descrito na Seção 3, também detalhado na Figura 2.

A abordagem apresentada mostrou-se viável ao apresentar média superior às soluções monolíticas. Além disso, percebeu-se que a nova proposta diminui consideravelmente a tendência observada no classificador monolítico. A adoção dos conceitos de SMC e Metalearning, implicou na distribuição mais adequada dos palpites como observada nas matrizes de confusão. Na análise de acurácia, a nova abordagem se apresenta superior tanto fold a fold como na média. A menor vantagem é observada na base BCC, mas evidentemente o aprendizado dos classificadores é prejudicado pelas poucas observações disponíveis para o estudo. Isso sugere que i) a técnica pode ser aplicada em outros contextos e, ii) pode ser fator limitante da técnica a baixa amostragem na base de dados.

## Referências

- Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer Publishing Company, Incorporated.
- Bashbaghi., S., Granger., E., Sabourin., R., and Bilodeau., G. (2017). Dynamic selection of exemplar-svms for watch-list screening through domain adaptation. In *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM*, pages 738–745. INSTICC, SciTePress. <https://doi.org/10.5220/0006256507380745>.
- Brazdil, P., Giraud-Carrier, C., Soares, C., and Vilalta, R. (2009). *Metalearning - Applications to Data Mining*. Springer.
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2):123–140. <https://tinyurl.com/ycr4yagn>.
- Brun, A. L., Britto, A. S., Oliveira, L. S., Enembreck, F., and Sabourin, R. (2018). A framework for dynamic classifier selection oriented by the classification problem difficulty. *Pattern Recognition*, 76:175 – 190. <https://doi.org/10.1016/j.patcog.2017.10.038>.
- Cruz, R., Sabourin, R., and Cavalcanti, G. (2018). Dynamic classifier selection: Recent advances and perspectives. *Information Fusion*, 41. <https://doi.org/10.1016/j.inffus.2017.09.010>.
- da Silva Borges, Z., Wehrmeister, F. C., Gomes, A. P., and Gonçalves, H. (2016). Exame clínico das mamas e mamografia: desigualdades nas regiões sul e nordeste do brasil. *Revista Brasileira de Epidemiologia [online]*, 19:1–13. <https://doi.org/10.1590/1980-5497201600010001>.
- Dua, D. and Graff, C. (2019). UCI machine learning repository. <http://archive.ics.uci.edu/ml>.
- Ferrari, D. G. (2014). *Seleção de algoritmos para a tarefa de agrupamento de dados: uma abordagem via meta-aprendizagem*. PhD thesis, Engenharia Elétrica, São Paulo, Brasil. <http://tede.mackenzie.br/jspui/handle/tede/1523>.
- Freund, Y. and Shapire, R. E. (1999). Large margin classification using the perceptron algorithm. *Machine Learning*, 37:277–296.

- Fusco, R., Sansone, M., Filice, S., Carone, G., Amato, D. M., Sansone, C., and Petrillo, A. (2016). Pattern recognition approaches for breast cancer dce-mri classification: A systematic review. *Journal of Medical and Biological Engineering*, 36:449–459. <https://doi.org/10.1007/s40846-016-0163-7>.
- IARC, I. A. f. R. o. C. (2020). Cancer today. <https://gco.iarc.fr/today/home>.
- Liu, X., Wang, X., and Matwin, S. (2018). Interpretable deep convolutional neural networks via meta-learning. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. <https://doi.org/10.1109/IJCNN.2018.8489172>.
- Mattos, J. S. C., Caleffi, M., and Vieira, R. A. C. (2013). Rastreamento mamográfico no brasil: Resultados preliminares. *Revista Brasileira de Mastologia*, 23(1):22–27. <https://tinyurl.com/y6w9c7cf>.
- Ministério da Saúde (2022). Prevenção do câncer feminino: Pesquisa do ministério da saúde revela aumento de exames de mamografia nos últimos 13 anos. Technical report.
- Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. The MIT Press, Massachusetts, MA, USA.
- Nagi, S. and Bhattacharyya, D. K. (2013). Classification of microarray cancer data using ensemble approach. *Network Modeling Analysis in Health Informatics and Bioinformatics*, 2:159–173. <https://doi.org/10.1007/s13721-013-0034-x>.
- Prudêncio, R. B. C. and Ludermir, T. B. (2007). Aprendizagem ativa para seleção de exemplos em meta-aprendizado. In *VI Encontro Nacional de Inteligência Artificial (ENIA)*, pages 1032–1041. <https://www.cin.ufpe.br/rbcp/papers/ENIA07a.pdf>.
- Rosenblatt, F. F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386–408. <https://tinyurl.com/y2dz6oph>.
- Silva, R. A. (2018). *Uso da centralidade de redes complexas na combinação e na seleção de classificadores*. PhD thesis, Pontifícia Universidade Católica do Paraná, Programa de Pós-graduação em Informática, Curitiba, Brasil. <https://tinyurl.com/y94p6bcq>.
- Silva, R. A., Britto Jr, A. d. S., Enembreck, F., Sabourin, R., and de Oliveira, L. E. S. (2020). CSBF: A static ensemble fusion method based on the centrality score of complex networks. *Computational Intelligence*, 36(2):522–556. <https://doi.org/10.1111/coin.12249>.
- Souza, B. F. (2010). *Meta-aprendizagem aplicada à classificação de dados de expressão gênica*. PhD thesis, Ciências da Computação e Matemática Computacional, São Carlos, Brasil. <https://tinyurl.com/y8e2wa4q>.
- Tabár, L., Dean, P. B., Chen, T. H.-H., Yen, A. M.-F., Chen, S. L.-S., Fann, J. C.-Y., Chiu, S. Y.-H., Ku, M. M.-S., Wu, W. Y.-Y., Hsu, C.-Y., Chen, Y.-C., Beckmann, K., Smith, R. A., and Duffy, S. W. (2019). The incidence of fatal breast cancer measures the increased effectiveness of therapy in women participating in mammography screening. *Cancer*, 125(4):515–523.
- Whalen, S. and Pandey, G. (2013). A comparative analysis of ensemble classifiers: Case studies in genomics. *The IEEE 13th International Conference on Data Mining (ICDM)*. <https://doi.org/10.1109/ICDM.2013.21>.