

Método FEATORA na classificação de câncer de mama utilizando imagens histopatológicas

FEATORA method in breast cancer classification from histopathological images

Leila Maria Vriesmann

Doutora em Informática, Universidade Estadual de Ponta Grossa,
Ponta Grossa, Paraná Brasil – lmvriesmann@uepg.br

Alceu de Souza Britto Jr.

Doutor em Informática, Universidade Estadual de Ponta Grossa,
Ponta Grossa, Paraná Brasil, Pontifícia Universidade Católica do Paraná, Curitiba, Paraná,
Brasil – alceubritto@gmail.com

Resumo

A classificação de tumores de mama em benignos ou malignos pode ser feita utilizando imagens histopatológicas. A seleção de áreas obviamente benignas por um sistema computacional (CAD - *computer-assisted diagnosis*) auxiliaria o trabalho do patologista. Após a extração de características das imagens, classificadores treinados diagnosticariam a presença ou não do câncer de mama. A seleção dinâmica de subconjunto de classificadores mais adequados para cada imagem poderia incrementar a taxa de reconhecimento dos tumores. Assim, o objetivo desse trabalho é verificar a taxa de reconhecimento de um método de seleção dinâmica de subconjunto de classificadores, o FEATORA, na classificação de câncer de mama em imagens histopatológicas. O método FEATORA utiliza intervalos de valores de atributos previsores contínuos, juntamente com taxas de probabilidade de reconhecimento de cada um dos classificadores. Os experimentos foram realizados nas imagens de ampliação de 200x. Notou-se que todos os esquemas do método FEATORA conseguiram melhores médias de taxa de reconhecimento do que os classificadores base. Quatro esquemas do FEATORA obtiveram também médias de taxa de reconhecimento maiores que o voto majoritário simples entre todos os classificadores base. Porém, os resultados (com pouco menos de 70% de taxa de reconhecimento) indicaram a inviabilidade do uso do método, com a configuração apresentada nesse trabalho, na prática do patologista. Aprimoramentos, mais estudos e experimentos são necessários.

Palavras-chave: Seleção Dinâmica de Subconjunto de Classificadores. Valores de Atributos Previsores. Imagens de Tumores de Mama.

Abstract

The classification of benign or malignant breast tumors can be done using histopathological images. The selection of obviously benign areas by a computer-assisted diagnosis (CAD) would aid the work of the pathologist. After the feature extraction from the images, trained classifiers would diagnose the presence or absence of breast cancer. Dynamic ensemble selection of classifiers for each image could increase the recognition rate of tumors. Thus, the objective of this study is to verify the recognition rate of a dynamic ensemble selection method of classifiers, the FEATORA, in the classification of breast cancer from histopathological images. The FEATORA method uses intervals of continuous predictive attribute values, together with recognition probability rates for each of the classifiers. The experiments were performed at 200x magnification. It was noted that all schemes of the FEATORA method achieved better average recognition rates than base classifiers. Four FEATORA schemes also obtained average recognition rates greater than the simple majority vote among all the base classifiers. However, the results (less than 70% of recognition rate)

indicated the non-viability of the method, with the configuration presented in this study, in the practice of the pathologist. Improvements, more studies and experiments are needed.

Keywords: Dynamics Ensemble Selection of Classifiers. Predictive Attribute Values. Breast Tumor Images.

INTRODUÇÃO

Anualmente, muitos casos de câncer de mama têm sido diagnosticados. Segundo Cazap (2018), só na América Latina, há o diagnóstico de aproximadamente 115.000 mulheres e 37.000 mortes em decorrência de câncer de mama por ano.

Um passo muito importante para o tratamento do câncer é identificação de tecidos normais, tecidos benignos e tecidos malignos (NAHID; KONG, 2017). Imagens histopatológicas têm sido examinadas intensivamente por médicos para descobrir a situação do paciente (NAHID; KONG, 2017). Se áreas obviamente benignas fossem filtradas por um diagnóstico assistido por computador (CAD - *computer-assisted diagnosis*), o trabalho dos patologistas seria aliviado, permitindo que se concentrassem em casos mais complexos, como apontado por GURCAN et al. (2009).

A classificação de imagens histopatológicas de tumores de mama em benignas ou malignas (câncer de mama) é tratada em vários trabalhos, como em Spanhol et al. (2016a), em Spanhol et al. (2016b) e em Gupta e Bhavsar (2017), onde a base de dados BreakHis (SPANHOL et al., 2016a) foi utilizada.

Em Spanhol et al. (2016a), a base de dados BreakHis é apresentada. As imagens passaram por processo de extração de características (cinco descritores de textura e um descritor de ponto-chave). Então, quatro classificadores distintos (1-NN – *1-nearest neighbor*, QDA – *quadratic linear analysis*, SVMs – *support vector machines*, RF – *random forests of decision trees*) foram treinados e testados. Médias de taxa de reconhecimento entre 80% e 85% na classificação, a nível de paciente, chegaram a ser atingidas.

Já em Spanhol et al. (2016b), a arquitetura CNN (*Convolutional Neural Networks*) (LECUN et al., 1989) AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) foi adaptada para a classificação. Para o treinamento da arquitetura, diversas estratégias foram propostas. Os experimentos foram avaliados a nível de paciente e a nível de imagem. CNN apresentou incrementos na taxa de reconhecimento, quando comparada a classificadores tradicionais treinados com descritores de textura.

Em Gupta e Bhavsar (2017), características de textura e cor (*colour-texture features*) obtidas das imagens histopatológicas foram concatenadas para treinar 22 classificadores. Houve experimento onde os classificadores foram treinados e testados utilizando imagens de todas as ampliações (sistema independente de ampliação), e experimentos onde era realizado o treinamento em uma ampliação específica (por exemplo, 40x) e o teste acontecia em diferentes ampliações (40x, 100x, 200x, 400x). Para cada experimento, a classe (benigna ou maligna) da imagem era obtida pelo voto majoritário entre todos os classificadores. As taxas de reconhecimento foram avaliadas a nível de paciente. No experimento independente de ampliação, o método proposto obteve melhores resultados do que uma abordagem baseada em aprendizagem profunda. Nos demais experimentos, os resultados mais consistentes ocorreram nas ampliações intermediárias (100x, 200x) e a média de precisão desses (média entre todas as ampliações) foi similar à média de precisão do experimento independente de ampliação (GUPTA; BHAVSAR, 2017).

Spanhol et al. (2016a) sugere que a combinação ou a seleção de classificadores poderia melhorar a taxa de reconhecimento na base de dados BreakHis (SPANHOL et al., 2016a). Um método para seleção dinâmica de subconjunto de classificadores (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) é o FEATORA, apresentado em Vriesmann e Britto Jr. (2015). O método FEATORA obteve resultados promissores, em Vriesmann e Britto Jr. (2015), no reconhecimento de variedades de trigo, em dados obtidos de imagens de grãos.

Assim sendo, o objetivo do presente trabalho é verificar o comportamento do método FEATORA, em termos de taxa de reconhecimento, na classificação de imagens histopatológicas de câncer de mama. As imagens de ampliação de 200x da base de dados BreakHis (SPANHOL et al., 2016a) serão utilizadas, uma vez que Gupta e Bhavsar (2017) relatou resultados consistentes nessa ampliação.

MATERIAIS E MÉTODOS

Base de dados BreakHis e extração de características

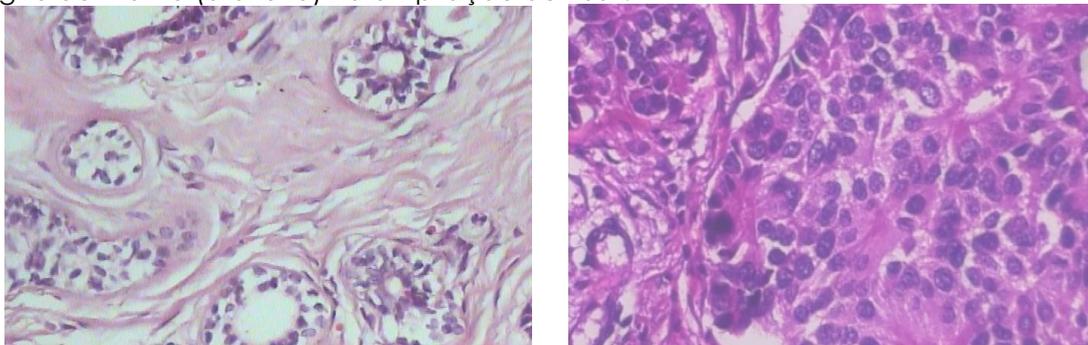
A base de dados BreakHis (SPANHOL et al., 2016a) possui imagens histopatológicas de tumores benignos e malignos (câncer) de mama nas ampliações de 40x, de 100x, de 200x e de 400x. Ao todo, são 7909 imagens, adquiridas de 82 pacientes.

Gupta e Bhavsar (2017), em seu estudo sobre a importância da ampliação na classificação de imagem histopatológica de câncer de mama, relatou resultados consistentes na ampliação de 100x e na ampliação de 200x. A média de precisão desses (média entre todos os fatores de ampliação) foi similar à média de precisão do experimento que utilizou todas as ampliações. O uso de imagens de apenas um fator de ampliação implica em menos tempo de processamento em relação ao uso de imagens de todos os fatores de ampliação, uma vez que, no primeiro caso, menos imagens seriam utilizadas. Assim sendo, optou-se por selecionar para esse trabalho apenas as imagens de ampliação de 200x.

Há 623 imagens de tumores benignos e 1390 imagens de tumores malignos na ampliação de 200x na base de dados BreakHis (SPANHOL et al., 2016a), totalizando 2013 imagens. A Figura 1 mostra na imagem da esquerda um tumor benigno e na imagem da direita um tumor maligno, ambas na

ampliação de 200x, adquiridas da base de dados BreakHis (SPANHOL et al., 2016a).

Figura 1: Imagem de um tumor benigno de mama (à esquerda) e imagem de um tumor maligno de mama (à direita) na ampliação de 200x.



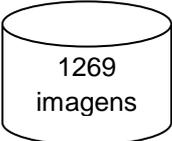
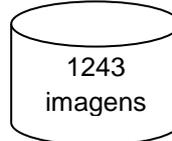
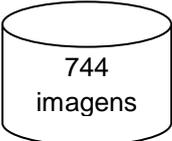
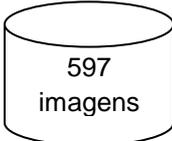
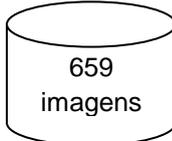
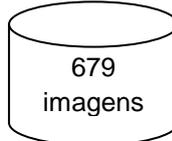
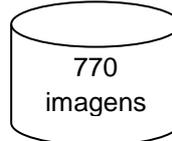
Fonte: Base de dados BreakHis (SPANHOL et al., 2016a).

Em Spanhol et al. (2016a), BreakHis foi dividida aleatoriamente em base de dados de treinamento (70%) e teste (30%), tomando cuidado para que as imagens dos pacientes utilizadas no treinamento não fossem também utilizadas na base de dados de teste. Foram geradas cinco divisões diferentes (*5-fold*), cada uma contendo uma base de dados de treinamento e uma de teste. A Figura 2 apresenta a quantidade de imagens selecionadas para cada uma das bases em cada divisão.

As divisões (ilustradas na Figura 2) foram obtidas, para o presente trabalho, após executar um código em Python (VAN ROSSUM, 2011), o qual está disponível em <http://www.inf.ufpr.br/lesoliveira/download/mkfold.tar.gz>. Depois, as imagens passaram pelo processo de extração de características, utilizando os pacotes *skimage* (VAN DER WALT et al., 2014) e *numpy* (OLIPHANT, 2006) em Python. Foram computadas as propriedades 'energy', 'homogeneity', 'contrast', 'dissimilarity', 'correlation', 'ASM' em quatro ângulos (0°, 45°, 90°, 135°) com 256 níveis de cinza de uma GLMC (*Gray-Level Co-Occurrence Matrix*). Assim, 24 atributos previsores contínuos foram obtidos para cada imagem. Juntamente com a classe (benigna ou maligna) da imagem, os 24 atributos previsores formaram uma instância, que podia ser de

treinamento, caso fosse da base de dados de treinamento, ou de teste, caso tivesse tido origem em uma das imagens de teste (Figura 3).

Figura 2: Quantidade de imagens na base de dados de treinamento e na base de dados de teste em cada uma das divisões de Spanhol et al. (2016a), obtidas utilizando o código em Python disponível na Internet.

1ª divisão	2ª divisão	3ª divisão	4ª divisão	5ª divisão
 <p>1269 imagens</p>	 <p>1416 imagens</p>	 <p>1354 imagens</p>	 <p>1334 imagens</p>	 <p>1243 imagens</p>
Base de dados de treinamento	Base de dados de treinamento			
 <p>744 imagens</p>	 <p>597 imagens</p>	 <p>659 imagens</p>	 <p>679 imagens</p>	 <p>770 imagens</p>
Base de dados de teste	Base de dados de teste			

Fonte: o autor, baseando-se nos resultados do código em Python obtido em <http://www.inf.ufpr.br/lesoliveira/download/mkfold.tar.gz>.

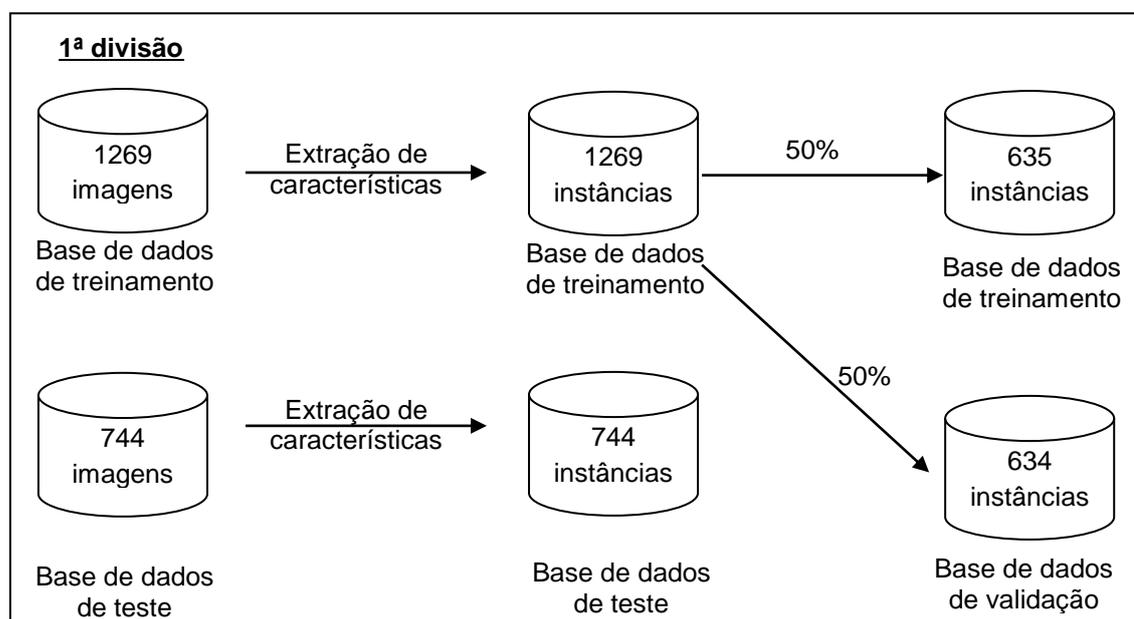
Em cada divisão representada na Figura 2, 50% dos dados de treinamento foram separados, a nível de imagens, para a base de dados de validação, restando apenas 50% para o treinamento. Quando a quantidade de instâncias era ímpar, a base de treinamento recebia uma instância a mais que a de validação. O processo foi realizado utilizando o filtro *StratifiedRemoveFolds* da ferramenta Weka (FRANK; HALL; WITTEN, 2016), e está ilustrado à direita na Figura 3.

Classificação e método FEATORA

Uma etapa importante no processo de classificação é a especificação dos classificadores que serão utilizados, uma vez que isso

influencia na taxa de reconhecimento das classes (benigna ou maligna) do tumor de mama. Como uma tentativa para melhorar a taxa de reconhecimento, em Spanhol et al. (2016a) houve a sugestão da combinação ou da seleção de classificadores. Um método para seleção dinâmica de subconjunto de classificadores (BRITTO JR.; SABOURIN; OLIVEIRA, 2014) é o FEATORA (VRIESMANN; BRITTO JR., 2015).

Figura 3: Representação do processo de extração de características e da divisão da base de dados de treinamento.



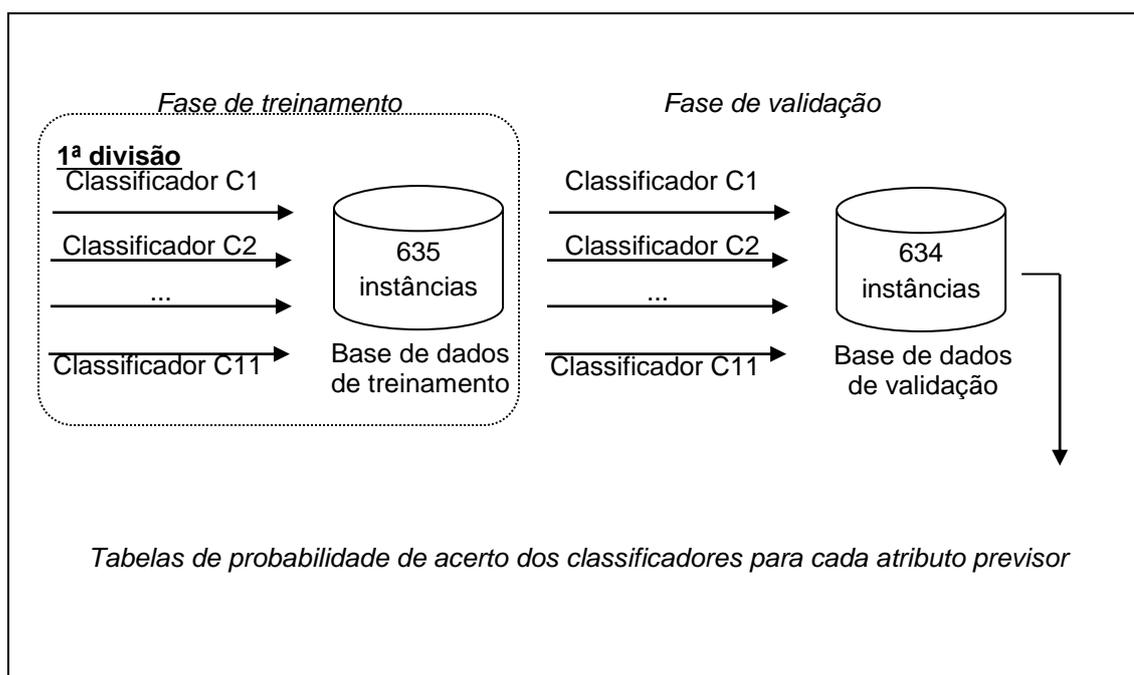
Fonte: o autor.

O método FEATORA necessita de uma base de dados de treinamento, uma base de dados de validação e uma base de dados de teste. Na primeira base de dados é realizado o treinamento dos classificadores. Nessa fase (quadro pontilhado da Figura 4), 11 classificadores distintos, disponíveis na ferramenta Weka (FRANK; HALL; WITTEN, 2016), foram executados, com configuração padrão da ferramenta nos parâmetros de entrada. Os classificadores utilizados foram: *BayesNet* (FRIEDMAN; GEIGER; GOLDSZMIDT, 1997), *DecisionTable* (KOHAVI, 1995), *Ibk* (AHA; KIBLER; ALBERT, 1991), *J48* (QUINLAN, 1993), *JRip* (COHEN, 1995), *Logistic* (LE CESSIE ; VAN

HOUWELINGEN, 1992), LWL (FRANK; HALL; PFAHRINGER, 2003; ATKESON; MOORE ; SCHAAL, 1997), *MultilayerPerceptron* (BISHOP, 1995), *OneR* (HOLTE, 1993), *RandomForest* (BREIMAN, 2001) e *REPTree* (ELOMAA; KAARIANEN, 2001).

Após a fase de treinamento, cada um dos classificadores utilizou a base de dados de validação para calcular sua probabilidade de reconhecimento em cada intervalo de valores de cada um dos atributos previsores (lado direito e parte inferior da Figura 4). Os intervalos de valores foram obtidos para cada atributo utilizando a fórmula de Sturges (STURGES, 1926). O processo foi codificado e executado em Octave (EATON et al., 2014).

Figura 4: Método FEATORA: fase de treinamento e fase de validação.



Fonte: o autor.

Atributo predictor 1					...	Atributo predictor 24					
Intervalo	C1	C2	...	C11		Intervalo	C1	C2	...	C11	
A1 ₁	┆	P1 _{1,1}	P1 _{1,2}	...	P1 _{1,11}	A24 ₁	┆	P24 _{1,1}	P24 _{1,2}	...	P24 _{1,11}
A1 ₂	┆	P1 _{2,1}	P1 _{2,2}	...	P1 _{2,11}	A24 ₂	┆	P24 _{2,1}	P24 _{2,2}	...	P24 _{2,11}
A1 ₃						A24 ₃					
...						...					

Na parte inferior da Figura 4, a Tabela “Atributo previsor 1” possui o primeiro intervalo de valores iniciando com o valor representado pela variável $A1_1$ e finalizando com o valor representado pela variável $A1_2$. Todas as instâncias de validação que possuem o primeiro atributo previsor nesse intervalo de valores foram classificadas (benignas ou malignas) pelo classificador $C1$. A probabilidade $P1_{1,1}$ foi obtida efetuando a operação matemática de divisão da quantidade de instâncias que o classificador $C1$ acertou a classe (benigna ou maligna), pela quantidade total de instâncias que tinham o valor desse atributo previsor no intervalo de $A1_1$ até $A1_2$. O mesmo procedimento foi repetido em todos os intervalos, em cada um dos 24 atributos previsores. Depois, o cálculo das probabilidades também foi feito para os demais classificadores ($C2, C3, C4, C5, C6, C7, C8, C9$ e $C10$).

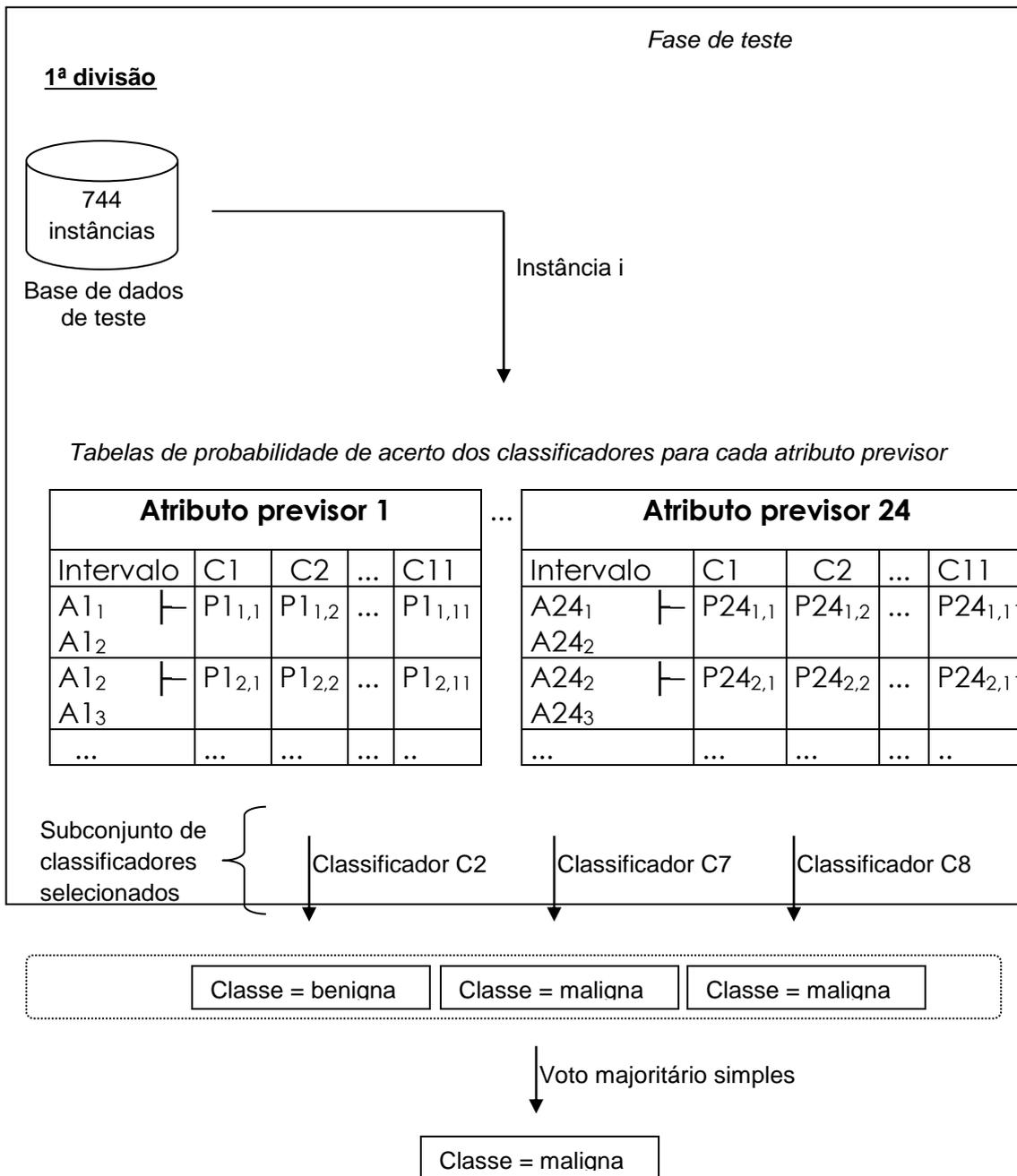
Após a criação das tabelas de probabilidade de acerto dos classificadores, em cada um dos atributos previsores, passou-se à fase de teste (Figura 5), executada em Octave (EATON et al., 2014). Um subconjunto de classificadores foi selecionado para cada instância de teste. Na Figura 5, como exemplo, foram selecionados os classificadores $C2, C7$ e $C8$, que forneceram as classes *benigna*, *maligna* e *maligna*, respectivamente, para a instância i . A classe final teve o valor *maligna*, pois foi a classe mais votada (voto majoritário simples). Para os experimentos, em caso de empate de votos, a classe de valor *benigna* era atribuída.

A seleção de cada um dos classificadores para compor o subconjunto considerou os valores dos atributos previsores da instância de teste. Para cada atributo previsor, é localizado o intervalo de seu valor na tabela, e é verificado se as probabilidades de reconhecimento dos classificadores atendem alguns critérios. Esses critérios variam de acordo com o esquema do FEATORA, mas levaram em consideração os parâmetros de entrada: um limiar superior de aceitação e, em alguns casos, um limiar inferior de eliminação. Para o primeiro parâmetro, foram selecionados os classificadores que tem probabilidade de reconhecimento maior ou igual ao limiar superior

de aceitação. No segundo parâmetro, foram excluídos os classificadores que têm probabilidade de reconhecimento menor que o limiar inferior de eliminação.

Em Vriesmann e Britto Jr. (2015), 14 esquemas diferentes foram propostos para o método FEATORA: FEATORA-ELIMINATE, FEATORA-ELIMINATE-W, FEATORA-UNION, FEATORA-UNION-W, FEATORA-UNION-U, FEATORA-UNION-ELIMINATE, FEATORA-UNION-ELIMINATE-W, FEATORA-UNION-ELIMINATE-U, FEATORA-UNIONL, FEATORA-UNIONL-W, FEATORA-UNIONL-U, FEATORA-ELIMINATEL, FEATORA-ELIMINATEL-W e FEATORA-MAJ. Além do uso dos limiares superior de aceitação e inferior de eliminação, os esquemas definem se os votos serão ponderados ou não. O voto majoritário simples entre os classificadores do subconjunto selecionado define a classe final da instância de teste. Cada um desses esquemas foi testado para a classificação das instâncias que referiam-se às imagens histopatológicas de câncer de mama.

Figura 5: Método FEATORA: fase de teste.



Fonte: o autor.

Métodos de comparação

Além do método FEATORA, outros métodos foram utilizados, para a comparação de resultados. Primeiramente, cada um dos 11 classificadores base (*BayesNet*, *DecisionTable*, *Ibk*, *J48*, *JRip*, *Logistic*, *LWL*, *MultilayerPerceptron*, *OneR*, *RandomForest* e *REPTree*) foi avaliado. Depois,

executou-se o voto majoritário simples entre todos os classificadores do conjunto inicial. Finalmente, foi calculado o valor do oráculo, ou seja, a taxa máxima de reconhecimento que seria possível com os classificadores do conjunto inicial. Em outras palavras, no oráculo supõe-se que, se há pelo menos um classificador no conjunto inicial que acerte a classe de uma instância de teste, haverá a classificação correta. O oráculo é um valor de taxa de reconhecimento que os métodos de seleção dinâmica de subconjunto de classificadores almejam atingir.

RESULTADOS E DISCUSSÃO

Os resultados foram avaliados utilizando a média (TR) e o desvio padrão (DP) das taxas de reconhecimento obtidas pelos métodos nas 5 divisões (*5-fold*). As taxas de reconhecimento foram avaliadas a nível de imagem, ou seja, a quantidade de instâncias de teste corretamente classificadas foi dividida pela quantidade total de instâncias de teste.

A Tabela 1 apresenta as TR e os DP obtidos com os classificadores base, com o voto majoritário simples e com o oráculo. O fato de serem utilizadas para o treinamento apenas metade das instâncias originalmente separadas por Spanhol et al. (2016a) para treinamento pode ter influenciado nas taxas de reconhecimento não terem atingido nem 70% (exceto no oráculo). Outro fator que pode ter influenciado diz respeito aos atributos previsores (propriedades de um GLMC), e a maneira como foram utilizados nas instâncias. Além disso, a configuração padrão dos parâmetros de entrada também poderia ser alterada, objetivando melhores resultados.

Na Tabela 1, *RandomForest* obteve a melhor média de taxa de reconhecimento (67,66%) entre os classificadores base, enquanto que *BayesNet* obteve a menor (61,24%). O voto majoritário simples obteve 69,27% de média de TR, sendo melhor que qualquer um dos classificadores base.

Nenhum dos métodos apresentados na Tabela 1 obteve o valor do oráculo, que foi de 91,34% de média de TR. Havia, então, uma margem de mais de 20% na taxa de reconhecimento que poderia ser incrementada, caso fosse utilizado um método de seleção dinâmica capaz de escolher o classificador (ou o subconjunto de classificadores) que acertasse a instância de teste.

O FEATORA, que é um método de seleção dinâmica de subconjunto de classificadores, foi aplicado. As execuções dos esquemas do método FEATORA iniciaram-se com os parâmetros de entrada em 50% no limiar inferior, e em 55% no limiar superior. Depois, foram incrementadas de cinco em cinco em cada nova execução, até atingir 95%, primeiramente no limiar superior, e depois no limiar inferior. Também foram feitos experimentos com limiar superior de 99%, e experimentos com limiar inferior de 99%.

Tabela 1: Médias de taxa de reconhecimento (e respectivos valores de desvio padrão) obtidos pelos métodos de comparação.

MÉTODO DE COMPARAÇÃO	TR(%)	DP
BayesNet	61,24	7,89
DecisionTable	67,59	2,72
IBk	61,33	2,87
J48	65,92	4,99
Jrip	66,25	7,08
Logistic	67,28	5,10
LWL	66,92	3,93
MultilayerPerceptron	66,00	5,72
OneR	62,08	2,46
RandomForest	67,66	4,21
REPTree	64,65	3,16
Voto Majoritário Simples	69,27	4,14
Oráculo	91,34	2,78

Tabela 1: Uma linha em negrito corresponde à maior média de taxa de reconhecimento (TR) entre os 11 classificadores base (11 primeiras linhas da tabela) e a outra linha em negrito corresponde à TR obtida com o voto majoritário simples entre os 11 classificadores base.

Fonte: o autor.

A Tabela 2 expõe a melhor média de taxa de reconhecimento (TR) de cada um dos esquemas do método FEATORA, juntamente com o valor do desvio padrão (DP), do limiar inferior de eliminação (quando for o caso) e do

limiar superior de aceitação da primeira das execuções que obteve a TR apresentada. Todas as TR apresentadas na Tabela 2 foram superiores às obtidas pelos 11 classificadores base (Tabela 1). Quatro esquemas do FEATORA obtiveram também médias de taxa de reconhecimento maiores que o voto majoritário simples (69,27%) da Tabela 1. Os valores de TR maiores obtidos foram 69,41% para FEATORA-UNION, 69,50% para FEATORA-UNION-W, 69,42% para FEATORA-UNION-U, e 69,39% para FEATORA-UNION-ELIMINATE-U.

O esquema FEATORA-UNION-W foi o que apresentou os melhores resultados (69,50% de TR) na Tabela 2. Nesse esquema, os classificadores que têm probabilidade de reconhecimento (calculada na base de dados de validação) com valor maior ou igual do que o limiar superior de aceitação são utilizados. O peso do voto de cada classificador selecionado é a sua probabilidade de reconhecimento.

No voto majoritário simples da Tabela 1, não houve empate, uma vez que a quantidade de classificadores era em número ímpar. Porém, na seleção dinâmica de classificadores pelo método FEATORA, empates poderiam acontecer. O fato de, em um empate, a classe atribuída ser benigna, pode ter influenciado negativamente na taxa de reconhecimento do FEATORA (Tabela 2), uma vez que a maioria das instâncias tinham classe com o valor maligna. O ideal seria estudar um tratamento diferente do proposto para o caso de empate.

O valor do oráculo (última linha da Tabela 1) não foi atingido por nenhum dos métodos avaliados, incluindo os esquemas do método FEATORA (Tabela 2).

Tabela 2: Melhores médias de taxa de reconhecimento (e respectivos valores de desvio padrão, de limiar inferior e de limiar superior) dos esquemas do método FEATORA.

ESQUEMA DO MÉTODO FEATORA	TR(%)	DP	LIMIAR INFERIOR(%)	LIMIAR SUPERIOR(%)
FEATORA-ELIMINATE	69,14	4,08	-	99
FEATORA-ELIMINATE-W	68,25	4,01	-	60
FEATORA-UNION	69,41	4,14	-	70

FEATORA-UNION-W	69,50	4,22	-	60
FEATORA-UNION-U	69,42	4,18	-	65
FEATORA-UNION-ELIMINATE	69,11	4,50	-	55
FEATORA-UNION-ELIMINATE-W	69,24	4,76	-	60
FEATORA-UNION-ELIMINATE-U	69,39	4,19	-	55
FEATORA-UNIONL	68,43	4,70	60	60
FEATORA-UNIONL-W	68,64	4,52	50	55
FEATORA-UNIONL-U	68,54	5,04	60	65
FEATORA-ELIMINATEL	68,48	4,36	50	99
FEATORA-ELIMINATEL-W	68,46	4,46	60	60
FEATORA-MAJ	68,95	4,64	60	65

Tabela 2: As linhas em negrito correspondem às médias de taxa de reconhecimento (TR) que foram maiores que 69,27% (do voto majoritário simples na Tabela 1).

Fonte: o autor.

O teste de Friedman, disponível em SciPy (JONES et al., 2001), foi aplicado nas taxas de reconhecimento obtidas nas divisões e, utilizando todos os métodos, mostrou que há diferenças. Como exemplo, foi testado apenas as TR de *BayesNet*, de *DecisionTable*, de *IBK* e de *J48*, e o teste revelou que a distribuição era parecida ($p > 0,05$). No entanto, houve diferenças tanto acrescentando o voto majoritário quanto acrescentando qualquer um dos esquemas do FEATORA, juntamente com esses quatro classificadores base. Não houve diferença quando as TR dos esquemas do FEATORA foram comparadas somente entre si (sem os classificadores base e sem o voto majoritário).

CONCLUSÃO

Um método de seleção dinâmica de classificadores, chamado FEATORA, foi aplicado na classificação de imagens histopatológicas de tumores da mama. Os tumores foram classificados em benignos ou malignos.

As imagens de ampliação de 200x da base de dados BreakHis (SPANHOL et al., 2016a) utilizadas passaram por um processo de extração de características, antes do uso do FEATORA. Todos os esquemas do método FEATORA conseguiram melhores de taxa de reconhecimento a nível de

imagens, do que os classificadores base. Quatro esquemas do FEATORA obtiveram também médias de taxa de reconhecimento maiores que o voto majoritário simples.

Embora a taxa de reconhecimento foi incrementada com o uso do método FEATORA, os valores obtidos foram abaixo do que se almejava. Disso, deduz-se a inviabilidade do uso do método, com a configuração apresentada nesse trabalho, na prática do patologista no diagnóstico de câncer de mama.

Porém, trabalhos futuros podem ser realizados para tentar melhorar a taxa de reconhecimento. O uso de outros métodos para extração de característica, da seleção de atributos, de outros métodos de classificação, de imagens de outras ampliações e a divisão da base de treinamento a nível de paciente são algumas sugestões de trabalhos futuros. Quanto ao método FEATORA, pode-se estudar aperfeiçoamentos, como um tratamento diferenciado no caso de empate nos votos das classes.

Mesmo não atingindo altas taxas de reconhecimento no problema em questão, o FEATORA mostrou-se promissor como método de seleção dinâmica de subconjunto de classificadores, obtendo incrementos em relação aos métodos comparados.

Esse foi um entre tantos trabalhos na literatura que têm sido realizados visando testar métodos e ideias no processo de classificação de imagens histopatológicas de câncer de mama. Espera-se que, no futuro, métodos computacionais sejam criados ou aperfeiçoados de modo a ficarem cada vez mais apropriados para os diagnósticos (como uma segunda opinião) na prática, aliviando o trabalho dos patologistas.

REFERÊNCIAS

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine Learning*, v. 6, p. 37-66, 1991.

ATKESON, C.; MOORE, A.; SCHAAL, S.. Locally weighted learning. *Artificial Intelligence Review*, v. 11, p. 11-73, 1997.

BISHOP, C. M. *Neural networks for pattern recognition*. New York: Oxford University Press, 1995.

BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.

BRITTO JR., A. S.; SABOURIN, R.; OLIVEIRA, L. E. S. Dynamic selection of classifiers – a comprehensive review. *Pattern Recognition*, v. 47, n. 11, p. 3665-3680, 2014.

CAZAP, E. Breast cancer in Latin America: a map of the disease in the region. *American Society of Clinical Oncology Educational Book*, n. 38, p. 451-456, 2018.

COHEN, W. W. Fast effective rule induction. In: Twelfth International Conference on Machine Learning, 115-123, 1995.

EATON, J. W.; BATEMAN, D.; HAUBERG, S.; WEHBRING, R. GNU Octave Version 3.8.1 Manual: a high-level interactive language for numerical computations. CreateSpace Independent Publishing Platform. 2014. Disponível em: < <http://www.gnu.org/software/octave/doc/interpreter> >. Acesso em: 10 out. 2018.

ELOMAA, T.; KAARIAINEN, M. An analysis of reduced error pruning. *Journal of Artificial Intelligence Research*, v. 15, p. 163-187, 2001.

FRANK, E.; HALL, M. A.; WITTEN, I. H. The WEKA workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

FRIEDMAN, N.; GEIGER, D.; GOLDSZMIDT, M. Bayesian network classifiers. *Machine Learning*, v. 27, p. 131-163, 1997.

GUPTA, V.; BHAVSAR, A.. Breast cancer histopathological image classification: is magnification important? In: 2017 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION WORKSHOPS (CVPRW), p. 769-776, 2017.

GURCAN, M. N.; BOUCHERON, L.; CAN, A.; MADABHUSHI, A.; RAJPOOT, N.; YENER, B. Histopathological image analysis: a review. *IEEE Reviews in Biomedical Engineering*, v. 2, p. 147-171, 2009.

HOLTE, R. C. Very simple classification rules perform well on most commonly used datasets. *Machine Learning*. v. 11, p. 63-90, 1993.

JONES, E; OLIPHANT, E; PETERSON P; et al. SciPy: Open Source Scientific Tools for Python, 2001. Disponível em: < <http://www.scipy.org/> >. Acesso em: 10 out. 2018

KOHAVI, R. The power of decision tables. In: 8th European Conference on Machine Learning, p. 174-189, 1995.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: 26th ANNUAL CONFERENCE ON NEURAL INFORMATION PROCESSING SYSTEMS (NIPS), 2012, P. L. Bartlett, F. C. N. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., Dec. 2012, p. 1106-1114. Disponível em: < <http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks> >. Acesso em: 10 out. 2018.

LE CESSIE, S.; VAN HOUWELINGEN, J. Ridge Estimators in Logistic Regression. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, v. 41, n. 1, p. 191-201, 1992.

LECUN, Y.; BOSER, B.; DENKER, J. S.; HENDERSON, D.; HOWARD, R. E.; HUBBARD, W.; JACKEL, L. D. Backpropagation applied to handwritten zip code recognition. *Neural Computation*, v. 1, n. 4, p. 541-551, 1989.

NAHID, A.; KONG, Y. Involvement of machine learning for breast cancer image classification: a survey. *Computational and Mathematical Methods in Medicine*, v. 17, p. 1-29, 2017.

OLIPHANT, T. E. Guide to NumPy. Trelgol Publishing, 2006. Disponível em: < <http://web.mit.edu/dvp/Public/numpybook.pdf> >. Acesso em: 10 out. 2018.

QUINLAN, R. C4.5: programs for machine learning. Morgan Kaufmann Publishers, San Mateo, CA, 1993.

SPANHOL, F. A.; OLIVEIRA, L. S.; PETITJEAN, C.; HEUTTE, L. A dataset for breast cancer histopathological image classification. *IEEE Transactions On Biomedical Engineering (TBME)*, v. 63, n. 7, p.1455-1462, 2016a.

SPANHOL, F. A.; OLIVEIRA, L. S.; PETITJEAN, C.; HEUTTE, L. Breast cancer histopathological image classification using convolutional neural networks. In: 2016 International Joint Conference on Neural Networks (IJCNN), p. 2560-2567, 2016b.

STURGES, H. A. The choice of a class interval. *Journal of the American Statistical Association*, v. 21, n. 153, p. 65-66, 1926.

VAN DER WALT, S.; SCHÖNBERGER, J. L.; NUNEZ-IGLESIAS, J.; BOULOGNE, F.; WARNER, J. D.; YAGER, N.; GOUILLART, E.; YU, T.; and the scikit-image contributors. scikit-image: image processing in Python. *Peer J.*, v. 2, p. 1-18, 2014.

VAN ROSSUM, G.; DRAKE, F. L. The Python language reference manual. Network Theory Ltd., 2011.

VRIESMANN, L. M.; BRITTO JR., A. S. Seleção dinâmica de subconjunto de classificadores para identificação de variedades de trigo utilizando probabilidade. In: X CONGRESSO BRASILEIRO DE AGROINFORMÁTICA (X SBIAGRO), 2015, Ponta Grossa, Paraná. Anais... 2015.

AGRADECIMENTOS

Agradecemos à colega Dra. Diolete Marcante Lati Cerutti pelo auxílio em outras atividades da Universidade Estadual de Ponta Grossa (PR), especialmente em Trabalho de Conclusão de Curso, o que possibilitou mais tempo para a primeira autora dedicar-se a essa pesquisa, aos colegas Dra. Elaine Margarete Guimarães e Dr. José Carlos Ferreira da Rocha, pelo incentivo na retomada dos experimentos, ao Me. Silvio Alexandre Porto, pela implementação do código em Python que coletava valores numéricos armazenados em arquivos, e os unia a símbolos contidos no nome do arquivo, e a todos que de alguma maneira ajudaram. Agradecemos também ao Albert Hung-Ren Ko, ao Robert Sabourin e ao Alceu de Souza Britto Jr. pela metodologia de trabalho do método KNORA, que motivou a elaboração do método FEATORA. Também tem-se um agradecimento especial à Revista Saúde e Desenvolvimento, e a seus colaboradores.