

# An Investigation of Jaya Optimization for non-Hodgkin Lymphoma Classification

Guilherme Freire Roberto,  
Marcelo Zanchetta do Nascimento  
Faculdade de Computação (FACOM)  
Universidade Federal de Uberlândia (UFU)  
Uberlândia-MG, Brasil  
{guilhermefreire,marcelo.zanchetta}@gmail.com

Thaína Aparecida Azevedo Tosta  
Centro de Matemática, Computação e Cognição (CMCC)  
Universidade Federal do ABC (UFABC)  
Santo André-SP, Brasil  
tosta.thaina@gmail.com

Leandro Alves Neves,  
Matheus Gonçalves Ribeiro  
Departamento de Ciências de Computação e Estatística (DCCE)  
Universidade Estadual Paulista (UNESP)  
São José do Rio Preto-SP, Brasil  
{neves.leandro,matheus.gonca.mgr}@gmail.com

Alessandro Santana Martins  
Instituto Federal do Triângulo Mineiro (IFTM)  
Ituiutaba-MG, Brasil  
alessandrosm2006@yahoo.com.br

**Abstract**—In this paper, an investigation of the Jaya Optimization algorithm is presented as a feature selection method. The algorithm is applied on fractal features calculated from three different classes of colored histological non-Hodgkin Lymphoma images. The AUC obtained after the application of the adapted Jaya algorithm was of 0.899, which represents an improvement of 0.039 on the AUC value in relation to the result obtained by considering all features. The performance was also compared with classical feature selection methods such as ReliefF and Gain Ratio. The results obtained by the Jaya algorithm were better than those in terms of AUC value.

## I. INTRODUÇÃO

A histopatologia é uma área de estudo da anatomia patológica e tem como objetivo a identificação de alterações das características estruturais dos tecidos. A análise ocorre via microscópio e permite auxiliar o diagnóstico de possíveis neoplasias presentes nos tecidos [1]. Esse processo de análise pode ser apoiado por sistemas computacionais baseados em técnicas de visão computacional, nomeados sistemas de apoio ao diagnóstico (do inglês, *computer aided diagnosis* - CAD) [2]. Um sistema CAD desenvolvido para análises de imagens histológicas pode ser dividido em diferentes etapas, tais como, a quantificação das regiões de interesse e análises desses resultados. O processo de análise das características pode ser executado para identificar, por exemplo, os valores mais importantes para distinguir e reconhecer padrões das estruturas com e sem alterações.

Para essa tarefa são aplicados algoritmos conhecidos, tais como *Chi-squared Attribute Evaluation* ( $\chi^2$ ), *ReliefF*, *Gain Ratio* e *Support Vector Machine Attribute Evaluation* (SVMAE). Esses métodos utilizam abordagens estatísticas baseadas em distância para selecionar os atributos mais relevantes [3] ou aplicar o classificador diretamente em subconjuntos de atributos, para avaliação dos mesmos, como é o caso do SVMAE.

Recentemente, algoritmos evolutivos têm sido adaptados em aplicações para seleção de atributos [4] e [5]. Métodos comumente aplicados em problemas de otimização, como o Algoritmo Genético (AG), o *Ant Colony Optimization* (ACO) e o *Grey Wolf Optimization* (GWO) foram aplicados na área de aprendizado de máquina como seletores de atributos [6], [7] e [8]. Esses algoritmos dependem de parâmetros específicos, configurados na inicialização. O estudo de quais valores são mais adequados para cada um desses parâmetros definidos a cada aplicação diferente é um problema custoso em termos de tempo. Porém, não podem ser ignorados e a escolha de valores não ideais pode comprometer os resultados finais.

Com base nesse problema, foi proposto em 2016, um algoritmo evolutivo sem nenhum parâmetro específico, que recebeu o nome de Jaya [9]. Os únicos valores que precisam ser ajustados nesse método são dados pelo tamanho da população e número de iterações a serem executadas. Deste modo, o objetivo deste trabalho consiste em investigar o desempenho do algoritmo evolutivo de otimização Jaya quando aplicado como método de seleção de atributos.

O método proposto neste trabalho empregou dimensão fractal (DF) e lacunaridade (LAC) para extração de características de imagens histológicas de linfomas. Em seguida, o algoritmo Jaya foi aplicado para identificar os valores mais relevantes. Finalmente, o método *Rotation Forest* foi utilizado para classificação dos descritores. A abordagem proposta foi testada em imagens histológicas de três diferentes classes de linfomas não-Hodgkin.

Na segunda seção deste texto são apresentados trabalhos relacionados ao tema que serviram como base para o método apresentado. A metodologia desenvolvida é descrita na Seção III. Os resultados obtidos são apresentados e discutidos na Seção IV. Finalmente, na Seção V são apresentadas as considerações e conclusões do estudo.

## II. TRABALHOS RELACIONADOS

Aplicação de métodos de seleção de atributos tem como objetivo melhorar o desempenho de classificadores por meio da redução do tempo necessário para a construção do modelo de aprendizado, e o aumento dos índices de classificação, como a acurácia [10]. As principais abordagens para seleção de atributos são: *wrapper*, embutida, filtrada e híbrida.

A abordagem *wrapper* consiste em aplicar algoritmos de aprendizado supervisionado para classificar diferentes subconjuntos de atributos [11]. A abordagem embutida consiste em aplicar o classificador diretamente no conjunto de atributos para selecionar os mais relevantes [12]. A abordagem filtrada consiste em aplicar métodos estatísticos para selecionar os atributos mais relevantes. O *ReliefF* é um método baseado na abordagem filtrada que aplica métodos estatísticos baseados em distância para seleção de atributos [3]. A abordagem híbrida é uma combinação entre as abordagens filtrada e *wrapper*. Nessa abordagem, é aplicado um método estatístico no conjunto de atributos para definir um *ranking* de relevância.

Uma nova abordagem baseada em algoritmos evolutivos tem sido o foco de pesquisas recentes como forma de aprimorar os algoritmos de seleção de atributos. Em um estudo proposto por [6], uma abordagem baseada em AG foi aplicada para eliminar atributos irrelevantes. O uso do AG em associação com o classificador *Rotation Forest* melhorou a acurácia da classificação de imagens de tumores malignos e benignos de mama de 96,78% para 99,48%.

Outros algoritmos evolutivos como o ACO e o GWO também são aplicados como métodos de seleção de atributos [7], [8]. Em [13], foi aplicado *Particle Swarm Optimization* (PSO) em associação com o classificador *Support Vector Machine* (SVM) para reduzir o conjunto de atributos obtidos a partir de imagens histológicas de tumores de fígados de 102 para 20. A acurácia obtida foi de 84,31%.

Entretanto, algoritmos evolutivos são dependentes de diversos parâmetros específicos que devem ser configurados na inicialização. Uma escolha inadequada dos valores desses parâmetros pode dificultar a obtenção de bons resultados de classificação. Em [9], os autores apresentam o algoritmo evolutivo *Jaya*, o qual foi proposto para solucionar esse problema de configuração de parâmetros. Vale ressaltar que nesse algoritmo as configurações de entrada são definidas apenas pelos parâmetros tamanho da população e o número de iterações. A adaptação do *Jaya* para método de seleção de atributos, conforme descrita em [14], é o foco deste trabalho. O algoritmo é descrito em detalhes na subseção III-C.

## III. METODOLOGIA

### A. Base de Imagens

O método proposto neste trabalho foi testado em um banco público de imagens de linfoma e leucemia, proveniente de estudos realizados por pesquisadores do *National Cancer Institute* e do *National Institute on Aging*, ambos localizados nos Estados Unidos [15] e [16]. Nesse banco para representar melhor a prática clínica, ao invés do ambiente rigidamente

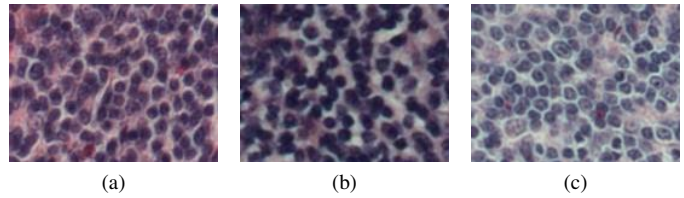


Figura 1. Imagens representantes das classes LCM (a), LF (b) e LLC (c)

controlado de laboratórios, as lâminas foram obtidas com significativas variações de corte e coloração. As imagens microscópicas foram adquiridas digitalmente, por meio de um microscópio de luz (*Zeiss Axioscope*) com objetiva de 20x e câmera digital colorida (*AXio Cam MR5*) acoplada. Regiões de interesse de cada lâmina foram selecionadas por especialistas, fotografadas digitalmente e gravadas sem compressão, padrão de cor RGB, resolução espacial de 1388x1040 *pixels*, com taxa de quantização de 24 *bits*. A base de imagens utilizada para os testes consiste em 173 imagens de Linfomas não-Hodgkin (LNH) sendo: 99 representantes do grupo Linfoma de Células do Manto (LCM), 62 representantes do grupo Linfoma Folicular (LF) e 12 representantes do grupo Leucemia Linfocítica Crônica (LLC). Na Figura 1 são apresentados um exemplo de cada classe.

### B. Extração de Características

A extração de características das imagens foi efetuada por meio da aplicação de técnicas baseadas na geometria fractal [17]. Esse conceito matemático foi apresentado, em 1975, por Mandelbrot para o estudo de formas que não podem ser definidas pela geometria euclidiana [18]. Esse modelo foi escolhido devido à presença de uma propriedade denominada auto-similaridade em imagens histológicas [19]. Isso pode ser visualizado ao observar-se uma mesma imagem sob escalas diferentes, onde é possível notar que as sub-imagens resultantes são similares entre si.

Em visão computacional, existem algumas abordagens para a observação de propriedades de auto-similaridade em imagens, denominadas multiescala. Entre as mais comuns, destacam-se o algoritmo *box-counting* e o *gliding-box*. Esses algoritmos têm como objetivo dividir as imagens em escalas diferentes e, a partir disso, extrair características de cada sub-imagem, gerando atributos locais para cada escala e globais que representam toda a imagem. Os atributos fractais selecionados foram a Dimensão Fractal (DF) e a Lacunaridade (LAC).

1) *Dimensão Fractal*: A obtenção da DF de uma imagem colorida, proposta em [20], foi calculada a partir de observações em diferentes escalas aplicando-se o algoritmo *gliding-box*. Este algoritmo consiste em posicionar em um canto da imagem uma caixa de tamanho  $L \times L$ . Essa caixa “desliza” pela imagem, *pixel a pixel*. Após percorrer toda a imagem, a caixa é reposicionada no ponto inicial e o valor de  $L$  é incrementado. Em uma imagem de tamanho  $H \times W$ , o

número de caixas  $T$  com lado  $L$  necessárias para percorrê-la por completo é apresentado na Equação 1:

$$T = (H - L + 1) \times (W - L + 1) \quad | \quad L \leq \min(H, W). \quad (1)$$

os valores de  $L$  variaram entre 3 e 45, com incremento de 2 unidades a cada iteração. A escolha desses valores foi feita devido à necessidade da existência de um *pixel* central na caixa.

Cada vez que a caixa é deslocada, o método proposto por Ivanovici e Richard [20] efetua uma análise de similaridade de cor em cada *pixel* da caixa. Essa análise é feita fixando-se o *pixel* central e atribuindo-o a um vetor  $F_c = (x_c, y_c, r_c, g_c, b_c)$ , onde  $x_c$  e  $y_c$  correspondem às coordenadas espaciais do *pixel* na imagem e  $r_c, g_c$  e  $b_c$  correspondem à intensidade luminosa de cada um dos canais de cor do modelo de cores RGB considerado neste estudo. Os demais *pixels* da caixa são atribuídos a um outro vetor  $F_i = (x_i, y_i, r_i, g_i, b_i)$  e esses são comparados ao *pixel* central por meio da distância de *Minkowski*, conforme exibido na Equação 2:

$$d = \max(|f_i - f_{c_i}|) \leq L, \forall i = \overline{1, 5}. \quad (2)$$

cada *pixel* que satisfaça  $d$  (ou seja, sua distância de *Minkowski* até o *pixel* central é menor ou igual que o tamanho da caixa  $L$ ) é considerado como pertencente a caixa.

Após calcular o número de *pixels* que satisfaçam  $d$  para cada caixa, e para diferentes valores de  $L$ , as informações obtidas foram usadas para a construção de uma estrutura conhecida como matriz de probabilidade. Cada elemento da matriz corresponde à probabilidade de que  $m$  *pixels* em uma escala  $L$  sejam rotulados como 1. Na Tabela I é apresentada uma representação visual desta matriz. É importante ressaltar que, apesar do nome, não se trata de uma simples matriz retangular pois enquanto seus valores das colunas cresçam linearmente, os rótulos das linhas são incrementados em potências de 2.

Tabela I  
REPRESENTAÇÃO DE UMA MATRIZ DE PROBABILIDADE.

	3	5	...	$L_{max}$
1	$P(1, 3)$	$P(1, 5)$	...	$P(1, L_{max})$
2	$P(2, 3)$	$P(2, 5)$	...	$P(2, L_{max})$
...	...	...	...	...
$L^2$	$P(L^2, 3)$	$P(L^2, 5)$	...	$P(L^2, L_{max})$

A matriz é normalizada de modo que a soma de todos os elementos de uma coluna seja igual a 1, conforme mostrado na Equação 3:

$$\sum_{m=1}^N P(m, L) = 1, \forall L. \quad (3)$$

Após a construção da matriz de probabilidade, os valores de DF locais são obtidos. Para cada tamanho de caixa  $L$ , a DF para aquela escala é denominada  $N(L)$ , a qual é calculada conforme mostrado na Equação 4:

$$N(L) = \sum_{m=1}^N \frac{P(m, L)}{m}. \quad (4)$$

Após o cálculo para todos os valores  $L$ , um gráfico que relaciona a evolução de  $\log L \times \log N(L)$  deve ser gerado. O valor do coeficiente angular da reta gerada corresponde ao valor global de DF. Como nem sempre uma reta perfeita é gerada, um método de regressão linear deve ser aplicado para a obtenção do atributo.

2) *Lacunaridade*: A LAC é uma medida complementar à DF que avalia de que forma o espaço de um fractal é preenchido (a DF avalia quanto o espaço é preenchido). Essa medida surge para lidar com situações em que imagens diferentes possuem os mesmos valores de DF.

Assim como na DF, foi proposto um método similar para a obtenção da lacunaridade de imagens coloridas [21]. Esse método é idêntico ao de obtenção da DF até o estágio de construção da matriz de probabilidade. Em seguida, foram calculados os momentos de primeira e segunda ordem a partir da matriz de probabilidade, conforme mostrado nas Equações 5 e 6:

$$\mu(L) = \sum_{m=1}^N m P(m, L), \quad (5)$$

$$\mu^2(L) = \sum_{m=1}^N m^2 P(m, L). \quad (6)$$

A lacunaridade em uma escala  $L$  é dada por  $\Lambda(L)$ , a qual é obtida de acordo com a Equação 7:

$$\Lambda(L) = \frac{\mu^2(L) - (\mu(L))^2}{(\mu(L))^2}. \quad (7)$$

Além de valores locais, neste trabalho foram avaliados atributos globais da lacunaridade, como explorado em [21] e [17]. Esses atributos são baseados em aplicações de métricas para extração de valores numéricos da curva de lacunaridade  $\log L \times \log \Lambda(L)$ , como área sob a curva, obliquidade, razão de área e escala do ponto máximo.

A área é calculada entre o eixo  $x$  e a função tomada como referência (no caso,  $\Lambda(L)$ ). Por definição, a área de uma função  $f$  entre dois pontos  $a$  e  $b$  sobre o eixo  $x$  pode ser obtida a partir da Equação 8:

$$A = \int_a^b f(x) dx. \quad (8)$$

Como as funções analisadas são discretas, a área é obtida por meio de integração numérica pelo método trapezoidal. Este método consiste em dividir a área em trapezoides com pontos igualmente espaçados entre si para um cálculo mais preciso da área. Para uma integração com pontos igualmente espaçados, a aproximação é dada pela Equação 9 em que o espaço entre cada ponto corresponde ao valor escalar:

$$\int_a^b f(x)dx \approx \frac{b-a}{2N} \sum_{n=a}^{b-1} (f(x_n) + f(x_{n+1})). \quad (9)$$

Na equação 7, as variáveis  $a$  e  $b$  correspondem aos valores do menor e do maior tamanho de  $L$ , respectivamente.

Outra métrica explorada foi a obliquidade, a qual deve ser entendida como uma medida da assimetria de uma amostra em relação à sua média. Se a obliquidade for negativa, a amostra está mais concentrada à esquerda do valor médio, se for positiva, está mais à direita. Em uma amostra perfeitamente simétrica, a obliquidade possui valor 0. Para uma amostra de  $N$  valores, a obliquidade  $O$  é dada pela Equação 10, sendo que  $\bar{x}$  corresponde à média da amostra e  $x_i$  corresponde ao valor da amostra referente ao  $i$ -ésimo valor de  $x$ :

$$O = \frac{\frac{1}{N} \sum_{i=a}^b (x_i - \bar{x})^3}{\sqrt{[\frac{1}{N-1} \sum_{i=a}^b (x_i - \bar{x})^2]^3}}. \quad (10)$$

A razão de área  $R$  também é outra métrica que pode ser extraída da função  $\Lambda(L)$ . A razão é calculada entre a área direita e a esquerda de cada função. Segundo os autores em [21], em texturas irregulares, esta razão se torna maior conforme a textura adquire mais heterogeneidade em maiores escalas. Dado que  $A_{a,b}$  corresponde à área formada pela curva da função entre os pontos  $a$  e  $b$  em relação ao eixo  $x$ , a razão de área é dada pela Equação 11:

$$R = \frac{A_{(\frac{b}{2}+1,b)}}{A_{(a,\frac{b}{2})}}. \quad (11)$$

A quarta métrica que pode ser extraída das funções é a observação da ocorrência do ponto máximo. Com isso, o vetor de características final é composto por 27 atributos: 1 atributo global de DF, 4 atributos globais de LAC e 22 atributos locais de LAC.

### C. Seleção de Atributos

Com o objetivo de reduzir o tempo necessário para o treinamento e a classificação das amostras, bem como remover atributos redundantes e irrelevantes, que podem dificultar a diferenciação entre as três classes de LNH, uma etapa de seleção de atributos foi agregada à metodologia.

O método escolhido foi o algoritmo evolutivo de otimização Jaya [9]. O Jaya difere de outros métodos evolutivos, como o algoritmo genético ou o *Ant Colony Optimization*, pela sua simplicidade e por não adotar nenhum parâmetro específico de inicialização. Seja  $\Gamma$  o número de soluções candidatas (tamanho da população) e  $J$  o total de atributos, a população inicial para a iteração  $i = 1$  é gerada, aleatoriamente, com base em uma distribuição uniforme no intervalo  $(0, 1)$  formando uma matriz  $\Gamma \times J$ . Entretanto, por ser um algoritmo inicialmente projetado para otimização de funções, o Jaya é implementado para um espaço contínuo de soluções. Para ser aplicado como um método de seleção de atributos, é necessária uma adaptação para um espaço binário (atributo  $j$  é selecionado ou não).

Para tal, a abordagem proposta em [14] nomeada pelos autores como *Binary Jaya Algorithm* (BJA) foi adotada.

No BJA, após a definição aleatória dos candidatos iniciais, os valores de seus atributos são normalizados conforme mostrado na Equação 12, sendo  $x_{i,j}^\gamma$  correspondente ao atributo  $j$  do candidato  $\gamma$  na iteração  $i$ :

$$x_{i,j}^\gamma = \begin{cases} 1, & \text{se } rand() > 0,5, \\ 0, & \text{caso contrário.} \end{cases} \quad (12)$$

Esses valores correspondem ao resultado da seleção de atributos. Caso  $x_{i,j}^\gamma$  seja 0, o atributo é removido. O vetor de características resultante para cada candidato é dado como entrada para o classificador. Em seguida, os candidatos são atualizados para a iteração seguinte de acordo com a Equação 13, em que  $M$  e  $P$  são os índices dos candidatos que obtiveram a melhor e a pior AUC, respectivamente, na iteração  $i$  e  $\alpha_{i,j}$  e  $\beta_{i,j}$  são dois valores gerados aleatoriamente no intervalo  $[0, 1]$  no início de cada iteração.

$$x_{i+1,j}^\gamma = x_{i,j}^\gamma + \alpha_{i,j}(x_{i,M}^\gamma - |x_{i,j}^\gamma|) - \beta_{i,j}(x_{i,P}^\gamma - |x_{i,j}^\gamma|). \quad (13)$$

Os valores obtidos após a aplicação da Equação 13 para cada atributo de cada candidato foram convertidos novamente para valores binários por meio do cálculo da tangente hiperbólica, conforme mostrado na Equação 14:

$$x_{i+1,j}^\gamma = \begin{cases} 1, & \text{se } rand() < \frac{e^{(|2x_{i,j}^\gamma|-1)}}{e^{(|2x_{i,j}^\gamma|+1)}}, \\ 0, & \text{caso contrário.} \end{cases} \quad (14)$$

Cada valor  $x_{i,j}^\gamma$  determina se um atributo  $j$  será ou não selecionado para aquele vetor de características candidato.

### D. Classificação e Avaliação de Desempenho

Para avaliar a capacidade de diferenciação entre as três classes de LNH, os vetores de características obtidos a partir da extração de atributos de DF e LAC foram dados como entrada para um classificador. Devido aos resultados relevantes apresentados em pesquisas recentes quando associado à métodos de seleção atributos evolutivos [6], o classificador escolhido foi o *Rotation Forest* com os seguintes parâmetros:

- 1) Classificador base: J48
- 2) Filtro de projeção: Análise de componentes principais
- 3) Instâncias removidas: 50%
- 4) Número de iterações: 10

Para o treinamento e teste das amostras foi adotada a estratégia de validação cruzada com *10-folds*. A plataforma *Weka v3.6.13* foi usada para a realização dos testes. A área sob a curva ROC (AUC) foi a métrica utilizada para medir o desempenho durante a classificação, servindo como parâmetro de otimização para o BJA. Outras medidas de desempenho foram verificadas para análise mais detalhada dos resultados: Taxa de Verdadeiros Positivos (VP), Taxa de Falsos Positivos (FP), Precisão (PREC), Sensitividade (SENS) e a Medida F (F-M).

Tabela II  
MATRIZ DE CONFUSÃO REFERENTE À CLASSIFICAÇÃO SEM SELEÇÃO DE ATRIBUTOS

LCM	LF	LLC	
83	13	3	LCM
13	47	2	LF
7	2	3	LLC

Tabela III  
RESULTADOS DETALHADOS DA CLASSIFICAÇÃO SEM SELEÇÃO DE ATRIBUTOS

VP	FP	PREC	SENS	F-M	
0,838	0,27	0,806	0,838	0,822	LCM
0,758	0,135	0,758	0,758	0,758	LF
0,25	0,031	0,375	0,250	0,300	LLC

#### IV. RESULTADOS E DISCUSSÃO

Os resultados apresentados nesta seção foram obtidos a partir de classificações das imagens de LNH com e sem a aplicação do BJA para seleção de atributos. Foi realizada em todos os testes uma comparação multiclases com as três classes de LNH utilizando o classificador *Rotation Forest*.

Os atributos DF e LAC foram extraídos das imagens histológicas de linfoma. Esses atributos foram usados para construção do vetor de características. Nessa primeira análise, os vetores foram empregados no classificador sem aplicação da etapa de seleção de atributos. A métrica AUC obtida foi de 0,860, com um total de 76,80% instâncias classificadas corretamente. A matriz de confusão resultante e outras métricas de avaliação são mostradas nas Tabelas II e III. Esses resultados indicam uma maior dificuldade na classificação da classe LLC, onde apenas 3 de um total de 12 instâncias foram classificadas corretamente.

O BJA foi aplicado para uma população de 20 candidatos. Após 10 iterações, a AUC obtida foi de 0,899, com uma acurácia de 78,03%. A matriz de confusão e os valores das demais métricas de avaliação são exibidos nas Tabelas IV e V.

É importante observar que a aplicação do BJA proporcionou uma melhora de 0,039 no valor de AUC. Houve também uma melhoria nas métricas referentes à classificação de LCM

Tabela IV  
MATRIZ DE CONFUSÃO REFERENTE À CLASSIFICAÇÃO COM SELEÇÃO DE ATRIBUTOS

LCM	LF	LLC	
88	9	2	LCM
14	45	3	LF
8	2	2	LLC

Tabela V  
RESULTADOS DETALHADOS DA CLASSIFICAÇÃO COM SELEÇÃO DE ATRIBUTOS

VP	FP	PREC	SENS	F-M	
0,889	0,297	0,800	0,889	0,842	LCM
0,726	0,099	0,804	0,726	0,763	LF
0,167	0,031	0,286	0,167	0,211	LLC

Tabela VI  
CONJUNTO DE ATRIBUTOS OBTIDOS COM INDICAÇÃO DOS SELECIONADOS PELO BJA

DF	$\Lambda(9)$	$\Lambda(17)$	$\Lambda(25)$	$\Lambda(33)$	$\Lambda(41)$	$O(\Lambda)$
$\Lambda(3)$	$\Lambda(11)$	$\Lambda(19)$	$\Lambda(27)$	$\Lambda(35)$	$\Lambda(43)$	$R(\Lambda)$
$\Lambda(5)$	$\Lambda(13)$	$\Lambda(21)$	$\Lambda(29)$	$\Lambda(37)$	$\Lambda(45)$	$Max(\Lambda)$
$\Lambda(7)$	$\Lambda(15)$	$\Lambda(23)$	$\Lambda(31)$	$\Lambda(39)$	$A(\Lambda)$	

(com exceção da PREC). Entretanto, o número de instâncias classificadas corretamente referentes às classes LF e LLC foi inferior quando comparado à classificação sem aplicação do BJA. Na Tabela VI, são indicados em negrito os atributos selecionados do vetor de características após a aplicação do BJA.

Para analisar o desempenho do BJA, foi feita uma comparação com métodos de seleção de atributos: SVMAE, *ReliefF*,  $Chi^2$  e *Gain Ratio*. Esses métodos foram escolhidos para comparação por serem representantes das abordagens filtrada e híbrida, as quais apresentam os resultados mais relevantes segundo informações publicadas em [10]. Os métodos foram aplicados sobre o mesmo conjunto de dados, gerados a partir da extração de características de DF e LAC das imagens de LNH. As métricas comparadas foram AUC, acurácia, número de atributos no vetor de características após a seleção e tempo de execução. Os resultados são exibidos na Tabela VII.

Tabela VII  
COMPARAÇÃO ENTRE DIFERENTES MÉTODOS DE SELEÇÃO DE ATRIBUTOS

Método	AUC	Acurácia	N. de atributos	Tempo (s)
Todos atributos	0,860	76,88%	27	0,00
SVMAE	0,895	81,50%	23	4.451,83
ReliefF	0,897	76,88%	14	0,03
$Chi^2$	0,892	80,92%	17	0,01
Gain Ratio	0,888	81,50%	21	0,01
BJA	0,899	78,03%	16	128,35

Todos os métodos obtiveram um desempenho próximo em termos de AUC e acurácia. Em relação ao número de atributos selecionados, o BJA e o *ReliefF* foram os métodos que geraram os menores vetores de características (16 e 14 atributos respectivamente), além de serem também os métodos que forneceram os melhores valores de AUC. Entretanto, o custo para a execução do BJA em termos de tempo é significativamente superior aos demais algoritmos, com exceção do SVMAE, conforme observado na Tabela VII. Isso ocorre pois tanto o SVMAE quanto o BJA dependem do resultado de classificações realizadas durante a execução do método, enquanto os demais algoritmos efetuam a seleção de atributos por meio de cálculos estatísticos sem o uso do classificador. Portanto, embora seja o método que obteve maior taxa de AUC, o BJA apresenta limitações em situações onde o tempo de execução é algo prioritário.

Para concluir a análise da eficiência do método proposto, uma comparação com resultados obtidos por trabalhos recentes na área de classificação de imagens de LNH foi feita. Os trabalhos selecionados foram [22], [23], [24] e [3], cujos

detalhes são exibidos na Tabela VIII. É possível notar que o método proposto neste trabalho apresentou desempenho inferior aos demais, o que indica que os descritores LAC e DF não são os ideais para este contexto. Entretanto, é importante ressaltar que o objetivo deste trabalho foi o de avaliar o desempenho do algoritmo de seleção de atributos BJA, e compará-lo com demais algoritmos de seleção disponíveis na literatura.

Tabela VIII  
COMPARAÇÃO COM TRABALHOS RECENTES RELACIONADOS À CLASSIFICAÇÃO DE IMAGENS DE LNH

Método	Descritores	Atributos	Acurácia
[22]	Wavelet, Gabor e LBP	50	92,70%
[23]	Atributos visuais	9.872	96,80%
[24]	Configurações de cor	216	95,50%
[17]	Percolação	3-8	92,00%
Proposto	DF e LAC	16	78,03%

## V. CONCLUSÃO

Neste trabalho foi apresentado o uso do algoritmo evolutivo Jaya como seletor de atributos para diferenciação entre três classes de LNH. O algoritmo foi aplicado em associação com as técnicas de extração de características DF e LAC. O método obteve uma AUC de 0,899 utilizando o classificador *Rotation Forest*, o que representa uma melhoria no valor de AUC de 0,039 quando comparado aos resultados obtidos sem o uso do seletor de atributos.

Embora a aplicação do algoritmo Jaya tenha acarretado em uma melhoria nos resultados, o custo da aplicação de um algoritmo evolutivo em termos de tempo é consideravelmente superior ao de métodos estatísticos tradicionais como o *ReliefF*. Em situações onde o número de iterações necessárias para a identificação dos atributos mais relevantes não é alto o suficiente para que o tempo de execução seja um problema, recomenda-se a aplicação do método proposto neste trabalho, por gerar resultados melhores que os demais em termos de AUC, além de reduzir consideravelmente o conjunto de atributos (sendo superado apenas pelo *ReliefF*).

Em trabalhos futuros pretende-se aplicar o BJA em vetores de características gerados a partir de outros métodos de extração. Além disso, é viável avaliar o comportamento do BJA quando associado à diferentes classificadores, como *Multilayer Perceptron*, SVM ou árvore de decisão.

## AGRADECIMENTOS

Os autores gostariam de agradecer à CAPES (32006012017P2), ao CNPQ (427114/2016-0) e à FAPEMIG (TEC-APQ-02885-15) pelo suporte financeiro.

## REFERÊNCIAS

[1] A. Bentaieb and G. Hamarneh, "Adversarial stain transfer for histopathology image analysis," *IEEE transactions on medical imaging*, vol. 37, no. 3, pp. 792–802, 2018.

[2] J.-Z. Cheng, D. Ni, Y.-H. Chou, J. Qin, C.-M. Tiu, Y.-C. Chang, C.-S. Huang, D. Shen, and C.-M. Chen, "Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in ct scans," *Scientific reports*, vol. 6, p. 24454, 2016.

[3] M. Robnik-Šikonja and I. Kononenko, "An adaptation of relief for attribute estimation in regression," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, vol. 5, 1997, pp. 296–304.

[4] M. Amoozegar and B. Minaei-Bidgoli, "Optimizing multi-objective pso based feature selection method using a feature elitism mechanism," *Expert Systems with Applications*, 2018.

[5] M. A. Al-Betar, M. A. Awadallah, H. Faris, I. Aljarah, and A. I. Hammouri, "Natural selection methods for grey wolf optimizer," *Expert Systems with Applications*, 2018.

[6] E. Aličković and A. Subasi, "Breast cancer diagnosis using ga feature selection and rotation forest," *Neural Computing and Applications*, vol. 28, no. 4, pp. 753–763, 2017.

[7] S. Tabakhi, P. Moradi, and F. Akhlaghian, "An unsupervised feature selection algorithm based on ant colony optimization," *Engineering Applications of Artificial Intelligence*, vol. 32, pp. 112–123, 2014.

[8] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, 2016.

[9] R. Rao, "Jaya: A simple and new optimization algorithm for solving constrained and unconstrained optimization problems," *International Journal of Industrial Engineering Computations*, vol. 7, no. 1, pp. 19–34, 2016.

[10] D. A. A. Gnana, S. Appavu, and E. J. Leavline, "Literature review on feature selection methods for high-dimensional data," *methods*, vol. 136, no. 1, 2016.

[11] Y. Zhang, S. Wang, P. Phillips, and G. Ji, "Binary pso with mutation operator for feature selection using decision tree applied to spam detection," *Knowledge-Based Systems*, vol. 64, pp. 22–31, 2014.

[12] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, "Gene selection for cancer classification using support vector machines," *Machine learning*, vol. 46, no. 1-3, pp. 389–422, 2002.

[13] W.-J. Kuo, "Computer-aided diagnosis for feature selection and classification of liver tumors in computed tomography images," in *2018 IEEE International Conference on Applied System Invention (ICASI)*. IEEE, 2018, pp. 1207–1210.

[14] T. Prakash, V. Singh, S. Singh, and S. Mohanty, "Binary jaya algorithm based optimal placement of phasor measurement units for power system observability," *Energy Conversion and Management*, vol. 140, pp. 34–35, 2017.

[15] N. C. I. NCI, <http://www.cancer.gov/>, 2018, acesso: 18/07/2018.

[16] N. I. on Aging NIA, <https://www.nia.nih.gov/>, 2018, acesso: 18/07/2018.

[17] G. F. Roberto, L. A. Neves, M. Z. Nascimento, T. A. Tosta, L. C. Longo, A. S. Martins, and P. R. Faria, "Features based on the percolation theory for quantification of non-hodgkin lymphomas," *Computers in biology and medicine*, vol. 91, pp. 135–147, 2017.

[18] B. B. Mandelbrot, "Les objets fractals: forme, hasard et dimension," 1975.

[19] L. Li, L. Chang, S. Ke, and D. Huang, "Multifractal analysis and lacunarity analysis: A promising method for the automated assessment of muskmelon (cucumis melo l.) epidermis netting," *Computers and electronics in agriculture*, vol. 88, pp. 72–84, 2012.

[20] M. Ivanovici, N. Richard, and H. Decean, "Fractal dimension and lacunarity of psoriatic lesions-a colour approach," *medicine*, vol. 6, no. 4, p. 7, 2009.

[21] A. Căliman and M. Ivanovici, "Psoriasis image analysis using color lacunarity," in *Optimization of Electrical and Electronic Equipment (OPTIM)*, 2012 13th International Conference on. IEEE, 2012, pp. 1401–1406.

[22] T. Meng, L. Lin, M.-L. Shyu, and S.-C. Chen, "Histology image classification using supervised classification and multimodal fusion," in *Multimedia (ISM)*, 2010 IEEE International Symposium on. IEEE, 2010, pp. 145–152.

[23] Y. Song, W. Cai, H. Huang, D. Feng, Y. Wang, and M. Chen, "Bioimage classification with subcategory discriminant transform of high dimensional visual descriptors," *BMC bioinformatics*, vol. 17, no. 1, p. 465, 2016.

[24] N. Codella, M. Moradi, M. Matasar, T. Sveda-Mahmood, and J. R. Smith, "Lymphoma diagnosis in histopathology using a multi-stage visual learning approach," in *Medical Imaging 2016: Digital Pathology*, vol. 9791. International Society for Optics and Photonics, 2016, p. 97910H.