

Uma Metodologia para Diagnóstico Automático da Filariose Utilizando Imagens Microscópicas Digitalizadas

LÚCIO FLÁVIO CAVALCANTI PESSOA¹
ASCENDINO FLÁVIO DIAS E SILVA²

Universidade Federal de Pernambuco
Centro de Tecnologia
Departamento de Eletrônica e Sistemas
Caixa Postal: 7800 Cidade Universitária
50732-970 Recife-PE Brasil
¹E-mail: L.PESSOA@IEEE.ORG
²E-mail: 13AFDS@NPD.UFPE.BR

Abstract. This paper describes a methodology for automatic diagnosis of filariasis, a tropical disease that represents a serious health problem in the State of Pernambuco, Northeast of Brazil. The medical diagnosis of filariasis is made scanning blood samples under an optical microscope and counting the number of a microscopic worm, commonly known as microfilariae. The methodology is based on the theories of Mathematical Morphology and Pattern Recognition and uses digital microscopic images, with resolution 640x480x64, as inputs. Using a training set with 56 patterns, the automatic recognition of microfilariae was performed by a Linear Discriminant Function with 4 features only, and the automation viability of this diagnosis was finally confirmed through the excellent classification results.

Keywords. Mathematical Morphology, Pattern Recognition, Image Analysis, Medical Imaging, Automatic Diagnosis of Filariasis

1. Motivação e descrição do problema

A filariose é uma doença tropical que representa um grave problema de saúde para o Estado de Pernambuco, Nordeste do Brasil. Recife (Capital do Estado) é considerada a capital mundial da filariose. Na tentativa de reverter esta situação lamentável, O Centro de Pesquisas Aggeu Magalhães (CPqAM), da Fundação Instituto Oswaldo Cruz (FIOCRUZ), vem desenvolvendo um sério programa de controle da doença.

A partir de levantamento estatístico nos bairros mais atingidos, entre 1989 e 1990, verificou-se uma incidência desta doença em até 17% da população dos respectivos bairros. Várias ações já foram tomadas e já vem sendo feito um controle do vetor da doença: *Culex Quinquefasciatus* (a muriçoca).

A filariose representa um grave problema social para nosso Estado, pois tem apresentado

uma maior incidência entre os jovens de 15 e 29 anos, que compõem a maior parte da população economicamente ativa.

Esta doença é causada pelo verme microscópico *Wuchereria Bancrofti* (microfilaria), que utiliza a muriçoca como agente transmissor, e produz deformações físicas em certas partes do corpo do ser humano. Os principais sintomas são fraqueza, perda de peso, febre intermitente e acesso de tosse noturna. Na fase aguda se manifesta com o aparecimento de ínguas na virilha e nos músculos das coxas, dores fortes e febre. Na fase crônica surge o inchaço muscular nos membros inferiores e superiores, testículos, mamas e vulva, que é vulgarmente chamado de elefantíase.

O diagnóstico médico da filariose é feito principalmente através da análise de lâminas de vidro, contendo amostras de sangue dos pacientes, em microscópicos ópticos. Após um

tratamento químico, estas lâminas são observadas exaustivamente por um técnico e, no final, é fornecido um diagnóstico através do número de microfilárias na amostra.

Este trabalho é bastante cansativo e sujeito a erros humanos, pois a lâmina precisa ser deslocada continuamente sob o campo visual do microscópico, uma vez que não é possível focar toda a amostra sangüínea de uma única vez.

Este processo de deslocamento, quando realizado manualmente, apresenta uma série de falhas. Por exemplo, podemos deixar de analisar uma certa região, ou contar uma certa microfilária mais de uma vez. Além disso, o exame é bastante desconfortável, podendo produzir problemas de vista no operador.

Com relação à precisão do exame, devemos salientar que o número de exames realizados por um certo operador, durante um dia, é um fator limitante na confiabilidade dos resultados. A medida que a carga de trabalho aumenta e muitos técnicos são convocados para desenvolver tarefas fora de sua área de especialização, a taxa de erro geralmente cresce, comprometendo a confiabilidade do exame.

Dessa forma, como o número de exames realizados por dia para detectar esta doença é grande, torna-se importante a tentativa de automatizar o seu diagnóstico, através de um equipamento que possibilite a emissão de um diagnóstico informatizado.

Em geral, além da minimização dos custos globais, aumento na confiabilidade dos resultados, e padronização dos exames, podemos caracterizar mais 3 benefícios básicos em utilizar um tal sistema automático para diagnóstico médico:

1. facilidade para detecção da doença em estágios preliminares, e numa ampla fração da população;
2. liberação dos laboratórios da análise cansativa do grande número de casos negativos;

3. aumento simultâneo em especificidade e sensibilidade do método de diagnóstico.

Neste sentido, foi desenvolvido um Sistema Automático de Leitura de Amostras em Microscópicos Ópticos [Silva (1992)] que representa uma proposta de automação nos exames realizados em microscópicos ópticos. Através deste sistema, é possível capturar (digitalizar) qualquer quadro de imagem em observação e proceder fazendo algum tipo de processamento ou análise digital de imagem. Este sistema está sendo utilizado para desenvolver um equipamento para diagnóstico automático da filariose.

De forma complementar, nosso trabalho tem sido no sentido de analisar o modelo matemático para identificação automática de microfilárias, e já foi verificada a viabilidade de automação do referido diagnóstico [Pessoa (1992)]. O presente trabalho apresenta um resumo de tais resultados.

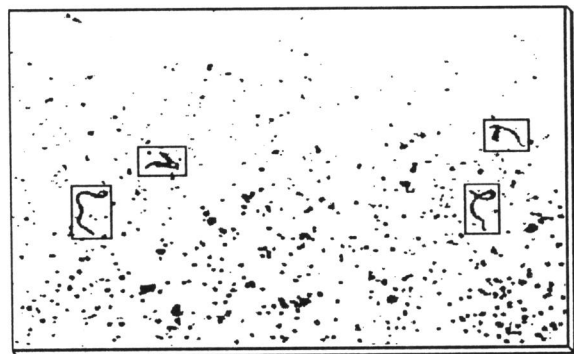


Figura 1: imagem contendo 4 microfilárias.

Para termos uma idéia da complexidade do problema, a Figura 1 apresenta um exemplo de uma imagem binária bastante característica para o exame da filariose, contendo 4 microfilárias e vários artefatos. As microfilárias são, em geral, as formas mais alongadas e estão destacadas através de retângulos. Os artefatos são apenas restos de materiais sangüíneos que

não foram totalmente retirados por ação química e, portanto, não têm nenhum interesse para diagnóstico.

2. Estratégia para solução do problema

O Sistema Automático de Leitura de Amostras em Microscópicos Ópticos, utilizado no presente trabalho, é composto de microscópio óptico, câmera CCD colorida, monitor de TV, controlador, teclado dedicado, digitizador de vídeo e microcomputador. O controlador é utilizado principalmente para acionar, independentemente do microcomputador, 3 motores de passo para as direções X-Y-Z. A movimentação X-Y possibilita varrer toda a lâmina de vidro com amostra sangüínea, e a movimentação Z é responsável pelo ajuste de foco.

Cada lâmina de vidro, contendo uma amostra sangüínea com cerca de 60µL, é composta de 3 esfregaços distintos do sangue de um paciente, aproximadamente com as mesmas dimensões. Por sua vez, cada esfregaço é sub-dividido imaginariamente em $M_i \times N_i$ quadros, com $i = 1,2,3$, onde cada quadro (campo) corresponde à área visível da amostra observada no microscópio. Finalmente, cada um desses quadros é analisado, fornecendo um diagnóstico parcial.

Devido ao processo ser essencialmente seqüencial, devemos ter a preocupação de interligar as informações de cada quadro, de tal maneira que os objetos que ficam na fronteira dos quadros sejam classificados corretamente. Os algoritmos ALG1 e ALG2 são utilizados para controlar o sistema e analisar automaticamente cada quadro, respectivamente.

Devemos salientar que o conhecimento da distribuição dos tamanhos (números de pixels) dos objetos, obtida após o passo 3 de ALG2 [Hoshen-Kopelman (1976)], é muito importante, pois já nos fornece uma primeira estimativa do número de microfilárias na amostra. De fato, como as microfilárias

apresentam um certo tamanho médio, aproximadamente da mesma ordem, é possível estabelecer uma faixa de variação ΔS_μ , tal que as microfilárias apresentem tamanhos numa faixa $S_\mu \pm \Delta S_\mu$. Dessa forma, verificando o tamanho do maior objeto ($S_{m\acute{a}x}$), se $S_{m\acute{a}x} \ll S_\mu - \Delta S_\mu$, podemos afirmar com certeza que não existem microfilárias no quadro em questão. Por outro lado, se $S_{m\acute{a}x} \gg S_\mu + \Delta S_\mu$, devemos ignorar o quadro e deixar a classificação para o operador, pois poderemos encontrar microfilárias superpostas com grandes artefatos, representando uma situação onde a análise automática não é feita corretamente. Finalmente, o número de objetos com tamanhos na faixa $S_\mu \pm \Delta S_\mu$ já representa a nossa primeira estimativa do número de microfilárias no quadro.

< ALG1: CONTROLE DO SISTEMA >
1. Inicializar esfregaço $i, i = 1, 2, 3 ; i \leftarrow 1 ;$
2. Digitalizar quadro $Q_{Kj}^i,$ $k = 1, \dots, M_i; j = 1, \dots, N_i ;$
3. Fazer análise do quadro Q_{Kj}^i e armazenar o resultado parcial;
4. Mover para o próximo quadro ;
5. Repetir os passos 2, 3 e 4 até o quadro $Q_{M_i N_i}^i ;$
6. $i \leftarrow i + 1 ;$ se $i \leq 3$ vá para 2 ;
7. Computar o número total de microfilárias em cada um dos 3 esfregaços de $M_i \times N_i$ quadros ;
8. Estimar o percentual de contaminação em função do número total de microfilárias na amostra e emitir diagnóstico final.

< ALG2: ANÁLISE DO QUADRO >
1. Binarizar o quadro utilizando 2 limiares ;
2. Aplicar o filtro morfológico ;
3. Aplicar o Algoritmo de Hoshen-Kopelman ;
4. Dependendo do valor de $S_{m\acute{a}x}$, analisar cada um dos objetos da imagem ;

5. Extrair os atributos pré-selecionados e classificar aqueles que não tocam o contorno do quadro ;

6. Caso o objeto toque o contorno do quadro, deixar a classificação pendente para quadros posteriores ;

Naturalmente, nesta primeira análise não foi considerado o problema dos objetos que tocam o contorno do quadro. Em geral, este é um problema difícil de ser resolvido, podendo ser inviável se o número de objetos que tocam o contorno for muito grande. Acreditamos que a eficiência na solução do mesmo depende da definição de uma adequada estrutura de dados. Esta representa uma das etapas do projeto, ainda em análise.

Quanto ao filtro morfológico [Haralick et al. (1987)] a ser utilizado, optamos pelo filtro passa-faixa (binário), no sentido de tamanho dos objetos. Sendo I a imagem original, A o elemento estruturante cruz 3×3 e T a imagem transformada, a definição do filtro passa-faixa é dada em (1).

$$T = T_1 \setminus T_2, \quad (1)$$

$$T_1 = \gamma(I; A), \quad (2)$$

$$T_2 = \varepsilon(T_1; rA), \quad (3)$$

onde γ é uma operação de abertura, ε é uma operação de erosão e r é o número de vezes que a operação de erosão é aplicada, ajustado para eliminar as microfílarias na imagem T_2 . Devemos lembrar que na imagem T_1 são eliminados os objetos de pequena amplitude e na imagem T , obtida pela subtração (binária) entre as imagens T_1 e T_2 , ficamos com os objetos de tamanhos numa certa faixa.

Para verificarmos a viabilidade de automação no diagnóstico da filariose, foi feito um estudo de casos utilizando 14 imagens digitalizadas com resolução $640 \times 480 \times 64$, contendo microfílarias e/ou artefatos.

Segundo análise visual destes 14 quadros, foram detectadas 22 microfílarias e um número

muito grande de artefatos. Foi feita então uma escolha aleatória de 2 artefatos em cada um dos 14 quadros, e em mais 3 quadros obtidos por uma operação de realce, fornecendo um total de $(14 + 3) \times 2 = 34$ artefatos. Dessa forma, foram obtidos $22 + 34 = 56$ objetos, a serem utilizados como Conjunto de Treinamento pictórico, apresentado na Figura 2, em nosso processo de análise. Cada uma das etapas deste processo serão descritas a seguir

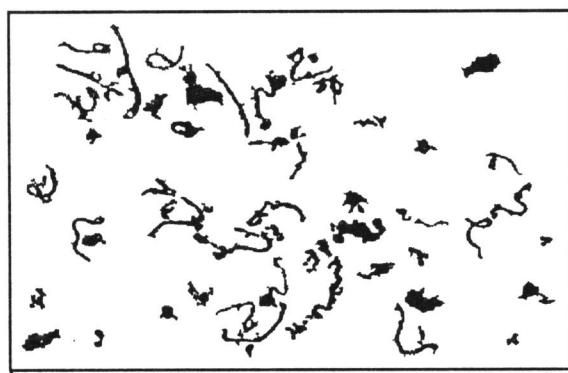


Figura 2: conjunto de treinamento pictórico.

2.1 Escolha dos limiares de binarização

A binarização das imagens foi realizada utilizando dois limiares de binarização, t_1 e t_2 , tal que todos os pixels com níveis de cinza entre t_1 e t_2 são transformados no nível lógico 1; caso contrário, no nível lógico 0. Após a binarização, cada pixel é então representado por apenas um bit.

Para obtermos os valores de t_1 e t_2 , procedemos determinando os histogramas de cada um dos 14 quadros e histogramas de janelas contendo somente microfílarias. Observando a média e o desvio padrão em cada um desses histogramas, ajustamos os limiares de binarização através de exaustiva análise visual das imagens binarizadas. Foram então verificados os seguintes valores médios de limiares:

$$t_1 = 7, \quad (4)$$

$$t_2 = 37. \quad (5)$$

O critério visual utilizado foi o de manter a correspondência entre as imagens original e binarizada, obtendo a melhor definição possível das microfilárias.

2.2 Aplicação do filtro morfológico

Conforme descrito anteriormente, utilizamos o filtro passa-faixa (1), com elemento estruturante cruz 3x3. De forma análoga a escolha dos limiares, após exaustiva análise nos 14 quadros disponíveis, tomando valores crescentes do parâmetro r, verificamos que o mesmo deveria ser ajustado para

$$r = 4. \tag{6}$$

2.3 Aplicação do algoritmo HK

O algoritmo de Hoshen-Kopelman (HK) foi introduzido inicialmente para resolver problemas de percolação, representando uma técnica largamente utilizada pelos Físicos para determinação da estatística de tamanhos de aglomerados (*clusters*). Na linguagem Física, um certo sistema é aproximado por uma rede, por exemplo a rede quadrada, e os seus elementos, denominados de sítios, podem ser de dois tipos: 0 e 1. Portanto, existe uma equivalência entre uma imagem binária e uma rede quadrada, onde os sítios do tipo 1 representam pixels de objetos da imagem.

Aproveitando esta equivalência, adaptamos o algoritmo HK para ser utilizado como uma ferramenta de análise de imagem, capaz de determinar a estatística de tamanhos (números de pixels) dos objetos de uma imagem binária.

A idéia de aplicar o algoritmo HK foi para ter uma estimativa do tamanho médio das microfilárias, bem como da diversidade (número de objetos diferentes) e de $S_{m\acute{a}x}$.

Foi feita então uma sub-divisão da classe de microfilárias em 3 sub-classes:

- 1) Microfilária Simples (MS);
- 2) Microfilária com Laço (ML);
- 3) Microfilária com Artefato (MA),

conforme ilustrado na Figura 3.

Após análise exaustiva em todas as janelas contendo microfilárias, nos 14 quadros, chegamos à primeira estimativa do tamanho médio das microfilárias em cada uma das sub-classes. Conforme podemos observar na Tabela 1, existe uma variação muito grande no tamanho típico dos objetos destas sub-classes, sendo insuficiente a utilização do tamanho como um critério de classificação.

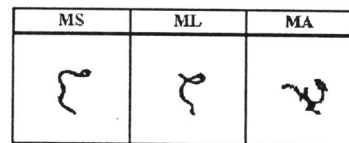


Figura 3: exemplos de microfilárias.

Microfilária	\bar{S} (pixels)
MS	$419 \pm 123,5 \sim 420$
ML	$302 \pm 103,1 \sim 300$
MA	$586 \pm 168,3 \sim 590$

Tabela 1: tamanhos médios de microfilárias.

2.4 Extração dos atributos

Objetivando uma análise detalhada das classes características do problema de reconhecimento automático das microfilárias, foram definidas m = 4 classes:

- Classe 1 : MS = C_1 ;
- Classe 2 : ML = C_2 ;
- Classe 3 : MA = C_3 ;
- Classe 4 : A = C_4 (Artefatos)

Após análise visual, foram então identificados n_i objetos em cada classe C_i , $i = 1, \dots, 4$. Observamos um total de $n_1 = 5$, $n_2 = 8$, $n_3 = 9$ e $n_4 = 34$. Observe que $n_1 + n_2 + n_3 = 22$, o número de microfírias observadas inicialmente.

Tendo em vista a aplicação da teoria de Reconhecimento de Padrões, obtivemos o Conjunto de Treinamento através da extração de 22 atributos, gerando um Espaço de Atributos com dimensão 22. Estes atributos são apresentados na Tabela 2.

Atributo	Descrição
X1	Área - A
X2	Massa - M Tamanho: número de pixels
X3	Perímetro - P
X4	Diâmetro equivalente 1 $(4A/\pi)^{1/2}$
X5	Diâmetro equivalente 2 $(6A/\pi)^{1/3}$
X6	Diâmetro equivalente 3 P/π
X7	Diâmetro de Feret 1 D_0
X8	Diâmetro de Feret 2 $D_{\pi/2}$
X9	Diâmetro de Feret 3 D_θ (corda máxima)
X10	Diâmetro de Feret 4 $D_{0+\pi/2}$
X11	Espessura $E = (1/4) [P - (P^2 - 16A)^{1/2}]$
X12	Comprimento $C = (1/4) [P + (P^2 - 16A)^{1/2}]$
X13	Orientação - θ direção da corda máxima
X14	Elongação 1 $D_0 / D_{\pi/2}$
X15	Elongação 2 $D_\theta / D_{0+\pi/2}$
X16	Elongação 3 C / E
X17	Fator de forma $F_f = 4\pi A/P^2$
X18	Momento central (1,1) μ_{11}

X19	Momento central (2,0) μ_{20}
X20	Momento central (0,2) μ_{02}
X21	Ângulo de invariância $\theta_m = (1/2) \arctg [2 \mu_{11} / (\mu_{20} - \mu_{02})]$
X22	Momento polar $\mu_p = \mu_{20} + \mu_{02}$

Tabela 2: relação dos atributos extraídos.

2.5 Estratégia de classificação

Após a aplicação do passo anterior, ficamos então com 56 vetores X de dimensão 22, denominados Padrões, dos quais sabemos a priori a classe que cada um deles pertence. Dessa forma, fica caracterizado um problema de Classificação de Padrões, analisado inicialmente com o enfoque estatístico, através de um processo de aprendizagem supervisionada.

Optamos pela aplicação de Funções Discriminantes Lineares (FDL's), usando conjuntamente as técnicas de Fisher [Fisher (1936)] e do Incremento Proporcional [Skansky-Wassel (1981)].

Para resolvermos o problema de m classes, utilizamos dicotomias de classes modificadas. Assim, definimos a classe modificada i (CM_i),

$$CM_i = CM_{i-1} \cup C_i, \quad i=2, \dots, m-1; \quad CM_1 = C_1, \quad (7)$$

e analisamos a dicotomia entre a classe modificada i e a sua complementar, denotada por dicotomia (CM_i / \overline{CM}_i).

A classificação é feita de forma sequencial, utilizando $m-1$ classificadores (funções discriminantes), onde $m = 4$ em nosso problema. Com efeito, seja X o Padrão que desejamos classificar, C o número da classe e $Z_i(X)$ o classificador correspondente à dicotomia (CM_i / \overline{CM}_i). O algoritmo de classificação para este tipo de dicotomia, denominado DICOT(m), é indicado no algoritmo ALG3.

< ALG3: DICOT(m) >
1. $i \leftarrow 1$;
2. Calcule $Z_i(\mathbf{X})$;
3. Se $Z_i(\mathbf{X}) \leq 0$, vá para 6 ;
4. $i \leftarrow i + 1$;
5. Se $i < m$, vá para 2 ;
6. $C \leftarrow i$;
7. Fim .

Observe que o algoritmo DICOT(m) tem como entrada o padrão \mathbf{X} , e como saída o número da classe que ele pertence.

A determinação dos classificadores Z_i segue alguns passos bem estabelecidos, indicados no algoritmo ALG4.

< ALG4: SÍNTESE DE Z_i >
1. Gere o Conjunto de Treinamento ;
2. Faça a seleção de atributos ;
3. Normalize o espaço ;
4. Calcule a Função Discriminante Z_i pela técnica de Fisher ;
5. Se a taxa de erro aparente for nula, vá para 7 ;
6. Ajuste Z_i através do Algoritmo do Incremento Proporcional ;
7. Fim .

Na tentativa de minimizarmos a dimensão do vetor padrão, além de aplicarmos uma técnica de seleção de atributos, decidimos também utilizar a análise *stepwise*. Assim, os passos 4 a 7 são repetidos utilizando uma dimensão crescente do vetor padrão, e no final é observada a taxa de erro aparente, obtida através do próprio Conjunto de Treinamento, em função da dimensão do Espaço. Dessa forma, determinamos a dimensão mínima ótima do vetor padrão de cada classificador Z_i .

A normalização do Espaço, indicada no item 3 de ALG4, transforma cada atributo X_i em X_i / σ_i , onde σ_i é o desvio padrão de X_i . Esta

etapa é indispensável, pois além de tornar o Espaço adimensional, proporciona uma redução de ruído.

2.6 Seleção de atributos

A seleção de atributos é uma etapa muito importante no processo de análise discriminante e anterior ao processo de aprendizagem, sendo responsável por uma redução de dimensionalidade do problema em análise.

Uma maneira bem simples e imediata de fazer uma seleção de atributos é através do critério de seleção de Fisher, usando o algoritmo proposto em [Pang et al. (1973)].

O critério de seleção de Fisher para um dado atributo é definido pela razão entre o valor absoluto da diferença das médias dos valores do atributo nas duas classes e a soma dos desvios padrões correspondentes. Ele é bastante eficaz, pois leva em consideração tanto informações interclasses como intraclasses.

Este critério fornece uma medida do erro de classificação que pode surgir quando o atributo correspondente é utilizado no classificador. Quanto maior for o critério de Fisher, menor será a probabilidade de erro de classificação, usando aquela variável.

Especificando o número máximo necessário de atributos ou o valor mínimo do critério de Fisher, podemos selecionar os atributos na ordem decrescente dos valores do tal critério, desprezando aqueles que trazem informações redundantes. A redundância é medida em função do coeficiente de correlação.

Em nosso problema, devemos fazer a seleção de atributos nas dicotomias (CM_i/\overline{CM}_i) , $i = 1,2,3$. Permitindo valores máximos do coeficiente de correlação de 0.7, 0.8 e 0.9, observamos que realmente houve uma redução de dimensionalidade gradativa, a medida que o coeficiente ficava menor.

Escolhemos então o caso correspondente ao menor coeficiente de correlação, pois nosso interesse é em utilizar o menor número possível de atributos. Coincidentemente, houve uma redução de 22 para 10 atributos e, a menos da ordem, os mesmos atributos foram selecionados para as 3 dicotomias em questão. Este resultado é bastante interessante, pois teríamos que extrair os mesmos atributos para os 3 casos.

A Tabela 3 apresenta os 10 atributos selecionados, na ordem decrescente de importância segundo o critério de Fisher. Devemos observar que o atributo de número 17, Fator de Forma, foi o mais discriminatório para as 3 dicotomias.

CM_1 / \overline{CM}_1	CM_2 / \overline{CM}_2	CM_3 / \overline{CM}_3
X17	X17	X17
X8	X11	X7
X11	X7	X10
X10	X10	X11
X7	X8	X8
X14	X15	X15
X18	X14	X18
X15	X21	X14
X13	X18	X13
X21	X13	X21

Tabela 3: relação dos atributos selecionados.

2.7 Análise *stepwise*

Aplicando a análise *stepwise* com os atributos da Tabela 3, determinamos 3 Funções Discriminantes Lineares de Fisher, para cada uma das dicotomias, com um número crescente de variáveis, na ordem da Tabela 3, e avaliamos a taxa de erro aparente. Finalmente, aplicamos o algoritmo do Incremento Proporcional e ajustamos os parâmetros das funções discriminantes para minimizar esta taxa de erro.

Verificamos que as taxas de erro aparente foram minimizadas em 8.9%, com dimensão 2,

21.4%, com dimensão 5, e 0.0%, com dimensão 4, nas dicotomias 1, 2 e 3, respectivamente.

Portanto, concluímos que não foi possível distinguir perfeitamente entre as classes C_1 , C_2 e C_3 (microfilárias), utilizando FDL's, mas que obtivemos pleno sucesso na separação entre as classes de microfilária e artefato, que era a nossa finalidade. Além disso, encontramos uma taxa de erro aparente nula em nosso Conjunto de Treinamento com 56 objetos, utilizando apenas 4 atributos.

A Tabela 4 resume estes 4 atributos do classificador da dicotomia (CM_3 / \overline{CM}_3), também na ordem de prioridade segundo o critério de Fisher. A expressão da respectiva FDL é indicada em (8), com $Y_i = x_{j_i} / \sigma_{j_i}$, $i = 1, \dots, 4$, $j_i = 17, 7, 10, 11$:

$$Z = w_0 + \mathbf{W}^T \mathbf{Y}, \quad (8)$$

onde,

$$w_0 = 1.250;$$

$$\mathbf{W} = (-0.4278, -0.5240, -0.3648, 0.6398)^T;$$

$$\mathbf{Y} = (Y_1, Y_2, Y_3, Y_4)^T.$$

Atributo	Descrição
X17 ($\sigma_{17} = 1.194$)	Fator de Forma $4\pi A/P^2$
X7 ($\sigma_7 = 0.8721$)	Diâmetro de Feret 1 D_0
X10 ($\sigma_{10} = 0.7717$)	Diâmetro de Feret 4 $D_{0+\pi/2}$
X11 ($\sigma_{11} = 0.5577$)	Espessura $(1/4) [P - (P^2 - 16A)^{1/2}]$

Tabela 4: atributos do classificador.

2.8 Teste do classificador

Esta é a última etapa de projeto do classificador. Devemos usar Padrões não pertencentes ao Conjunto de Treinamento, que definem o Conjunto de Teste, testar o

classificador original (8), e finalmente ajustar a FDL com a união do Conjunto de Treinamento e o Conjunto de Teste. Este ajuste tem como objetivo reduzir a dependência (vício) com o Conjunto de Treinamento.

Admitindo que o nosso Conjunto de Treinamento é representativo, optamos pela Geração Aleatória de Padrões [Pessoa (1990)]. Assim, geramos 2 novos Padrões dentro de uma hipersfera de raio ϵ , centrada em cada Padrão original, fornecendo um total de $2 \times 56 = 112$ Padrões no Conjunto de Teste.

Observando o mapeamento de (8) no Conjunto de Treinamento, e tomando valores de $\epsilon \sim |Z_{\min}|$, observamos o desempenho do classificador (8) após a perturbação, obtendo uma idéia de sua confiabilidade. A Tabela 5 apresenta as taxas de erro aparente do classificador (8) para 3 valores de ϵ , e os mesmos resultados após os ajustes, pelas técnicas de Fisher e do Incremento Proporcional, com a união dos dois conjuntos (Treinamento + Teste). A expressão final da FDL ajustada é indicada em (9):

$$Z_A = w_{0A} + W_A^T Y, \tag{9}$$

onde,
 $w_{0A} = 1.213$;
 $W_A = (-0.3650, -0.5264, -0.3933, 0.6596)^T$;
 $Y = (Y_1, Y_2, Y_3, Y_4)^T$.

ϵ	Teste	Treinamento + Teste
0.11	0.0%	0.0%
0.12	0.6%	0.0%
0.13	1.8%	1.2%

Tabela 5: taxas de erro aparente nos testes.

Observando a Tabela 5, podemos admitir que o classificador (9) terá desempenho perfeito em qualquer Padrão obtido de uma perturbação com $\epsilon = 0.12$. Naturalmente, para termos maior segurança neste resultado,

devemos utilizar o maior número possível de Padrões aleatórios. Este seria o classificador a ser utilizado no Sistema Automático de Leitura de Amostras em Microscópicos Ópticos.

2.9 Aplicação do algoritmo GL

Após a obtenção dos excelentes resultados de classificação anteriores, julgamos ter sido adequada a nossa definição das classes.

Visando constatar os nossos resultados e ter uma maior segurança, aplicamos o Algoritmo de Gitman-Levine (GL) [Gitman-Levine (1970)], uma técnica de agrupamento utilizando lógica difusa, e avaliamos o número de classes que foram geradas em função de um parâmetro característico do algoritmo.

Sem entrarmos em detalhes, verificamos que, em média, o número de classes do nosso problema foi em torno de 4, confirmando a nossa suposição inicial. Além disso, o número mínimo de classes geradas foi 2, confirmando que realmente é possível caracterizar, no mínimo, dois tipos de estruturas : microfilárias e artefatos.

3. Avaliação dos resultados

Apesar da grande diversidade das imagens utilizadas em nosso estudo de casos, concluímos que foi possível fazer o reconhecimento automático das microfilárias com uma taxa de erro aparente nula, sugerindo a viabilidade na automação deste diagnóstico.

Finalmente, após a aplicação do algoritmo de Gitman-Levine, constatamos a nossa escolha das classes, pelo menos em número de classes.

Verificamos também que os classificadores lineares não foram suficientes para resolver o problema das sub-classes de microfilárias. No entanto, o estudo com classificadores de segunda ordem pode ser completamente equivalente ao desenvolvimento apresentado neste trabalho, bastando trabalhar num espaço de dimensão maior. Apesar disso, nossa meta

de distinguir entre microfilárias e artefatos foi alcançada.

Como observação final, gostaríamos de salientar que as probabilidades a priori das classes de microfilária e artefato não são iguais, e talvez este fato possa ser utilizado para fazer polarização nos classificadores e ajustes necessários:

Agradecimentos

Gostaríamos de agradecer ao Centro de Pesquisas Aggeu Magalhães por parte do suporte financeiro ao presente trabalho, além da confiança depositada no Grupo de Pesquisa em Reconhecimento de Padrões da UFPE.

4. Referências

- [Fisher (1936)] R. A. Fisher, The use of multiple measurements in taxonomic problems. *Ann. Eugenics, Australia*, v.7, p. 179-188, 1936.
- [Gitman-Levine (1970)] I. Gitman, M. Levine, An algorithm for detecting unimodal fuzzy sets and its application as clustering technique. *IEEE Tans. on Comp.*, v. C-19, n.7, p.583-593, july 1970.
- [Haralick et al. (1987)] R. M. Haralick, S. R. Sternberg, X. Shuang, Image analysis using mathematical morphology. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, v.9, n.4, p. 532-550, july 1987.
- [Hoshen-Kopelman (1976)] J. Hoshen, R. Kopelman, Percolation and cluster distribution. I. Cluster multiple labeling technique and critical concentration algorithm. *Physical Review B*, v.14, n.8, p. 3438-3445, october 1976.
- [Pang et al. (1973)] C. K. Pang, A. J. Koivo, A. El Abiad, Application of pattern recognition to steady-state security evaluation in a power system. *IEEE Trans. Syst. Man and Cyb.*, v. SMC-3, n.6, p.449-458, 1973.
- [Pessoa (1990)] L. F. C. Pessoa, Gerador de padrões: uma alternativa na avaliação da segurança de sistemas de potência em regime permanente. Minas Gerais, 1990. 22p. (Prêmio CEMIG de Tecnologia, Categoria Estudantil, Nível Nacional).
- [Pessoa (1992)] L. F. C. Pessoa, Uma metodologia para diagnóstico automático da filariose utilizando imagens microscópicas digitalizadas. Recife, 1992. 144p. (Dissertação de Mestrado. UFPE).
- [Silva (1992)] A. H. S. C. Silva, Sistema automático de leitura de amostras em microscópicos ópticos. Recife, 1992. 183p. (Dissertação de Mestrado. UFPE).
- [Sklansky-Wassel (1981)] J. Sklansky, G. N. Wassel, *Pattern classifiers and trainable machines*. 1ª ed. USA: Springer-Verlag New York Inc., 1981.