

## Classificação de células de FAN Hep2

Leandro L. G. Oliveira<sup>1</sup>, Eduardo S. Albuquerque<sup>1</sup>, Paulo L.C. Francescantonio<sup>1</sup>

Departamento de Computação, LAPIMED - Universidade Católica de Goiás.  
Av. Universitária 1440 CEP74605-210, Goiânia, Goiás, Brasil  
{leandro, eduardo, paulo}@lapimed.net

**Abstract.** *This article shows preliminary results of a novel classification model for FAN Hep-2 exams. This work used wavelets transforms to extract features and Euclidian distance metrics to measure similarity between two images. The results obtained are promising. The work defines a standardized reading for FAN Hep2 Exams that can lead to a reduction of interpretation errors and can also be used to standardize FAN Hep-2 exam readings.*

**Resumo.** *Este artigo mostra os resultados preliminares de um novo modelo de classificação de exames de FAN Hep-2. As técnicas utilizadas neste artigo foram transformadas wavelets para extração de características e a métrica Euclidiana para medir a similaridade entre as imagens. Os resultados se mostraram promissores. O trabalho define uma leitura padronizada dos padrões, que poderá levar à redução dos erros de interpretação e poderá também ser utilizada para uniformizar a leitura dos exames FAN Hep-2.*

### 1. Introdução

Atualmente, as doenças auto-imunes estão incluídas entre as principais causas de problemas de saúde observados tanto em países industrializados quanto naqueles em desenvolvimento. Tem-se observado que, em estudos conduzidos em outros países, aproximadamente metade dos pacientes do sexo masculino portadores de artrite reumatóide está desempregada por causa de incapacidade física [1]. No Brasil, aproximadamente 10% da população são portadores de alguma doença auto-imune que pode levar à incapacidade física [2].

As doenças reumáticas auto-imunes, que são doenças crônico-degenerativas, estão incluídas no grande grupo das doenças auto-imunes e são caracterizadas por alterações na unidade morfofuncional do aparelho locomotor, com envolvimento de articulações e musculatura. A imunofluorescência indireta para pesquisa de anticorpos antinucleares - anticitoplasmáticos é uma técnica útil para se detectar marcadores de auto-imunidade não só em pacientes reumáticos, mas também, em outras doenças. O substrato para identificação de anticorpos antinucleares - anticitoplasmáticos foi durante muito tempo tecido de roedores, mas nos últimos 12 anos, foi substituído largamente por células HEp-2 [3]. Estas células imortalizadas originam-se de carcinoma laríngeo humano e crescem em monocamadas sobre lâminas de vidro [4]. As células HEp-2 permitem o reconhecimento de mais de 30 diferentes padrões nucleares e citoplasmáticos que são dados por mais de 100 diferentes auto-anticorpos. [4].

Apesar do FAN HEp-2 ser um método sensível, com capacidade de rastrear uma grande variedade de auto-anticorpos conhecidos e desconhecidos e oferecer indicações sobre a provável identidade do antígeno apresenta como limitação variações na

interpretação dos padrões de imunofluorescência. Diante da subjetividade na leitura dos diferentes padrões de imunofluorescência e da intensidade desta, pela utilização da visão e da memória, resultando em variações na interpretação, torna-se extremamente desejável a análise automática das imagens, processo este que pode ser desempenhado pelo computador com sucesso.

## 2. Modelo proposto

Este trabalho propõe uma metodologia de classificação de células de FAN Hep-2. A metodologia se divide em dois módulos que são: 1 - Formação do banco de conhecimento com imagens baseadas em laudos conhecidos (padrão ouro) e 2 – Classificação de novas imagens e sugestão do laudo. O processo de extração de característica é mostrado na seção 2.1. Utilizou-se para os testes os padrões nucleares (membrana nuclear, homogêneo, pontilhado pleomórfico, pontilhado do tipo pontos isolados, pontilhado grosso, pontilhado grosso reticulado, pontilhado fino, pontilhado centromérico, pontilhado fino denso) e nucleolares (nucleolar homogêneo, nucleolar aglomerado, nucleolar pontilhado). Para o primeiro módulo (formação do banco de conhecimento) utilizou-se 10 imagens para cada tipo de padrão totalizando 120 imagens. O segundo módulo (classificação de novas imagens) utilizou-se 20 novas imagens (total de 240 imagens) de cada padrão contido no banco de conhecimento. Portanto neste trabalho foram processadas 360 imagens contemplando 12 padrões de células de FAN Hep-2.

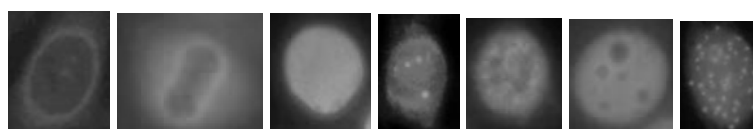


Figura 1: Exemplo de alguns padrões de imagens de FAN Hep-2.

Para seleção e extração das características utilizou-se as transformadas wavelets [5, 6]. As seguintes wavelets foram escolhidas para realizar os experimentos, Haar, Doubechies (Db2, Db4 e Db8). As transformadas wavelets decompõem uma imagem em um conjunto de coeficientes de aproximações ( $A_i$ ) e detalhes ( $D_i$ ) no nível  $i$  de decomposição. Este processo é repetido, com sucessivas decomposições a partir da aproximação anterior, gerando coeficientes de alta frequência ( $D_i$ ).

### 2.1. Cálculo do vetor de característica

Para a composição do vetor de característica, efetuou-se análise estatística (energia) [7] em cada um dos 3 níveis de coeficientes de detalhes ( $D_i$ ). O processo de composição do vetor de característica segue o esquema descrito na Figura 2. A energia foi computada pelo somatório dos quadrados das probabilidades de ocorrência de cada tom de cinza em cada nível de decomposição na imagem do padrão celular.

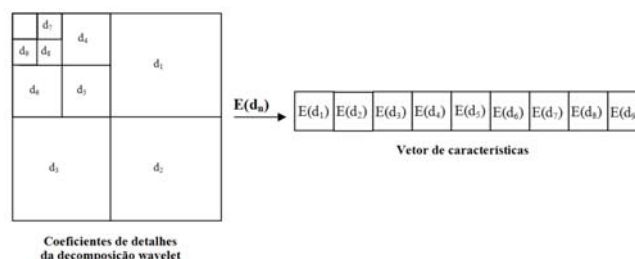


Figura 2: Composição do vetor de característica de uma imagem.

Várias medidas de similaridade, baseadas em estimativas empíricas da distribuição das características são utilizadas para diferentes tarefas da área de visão computacional. Os vetores de característica definidos neste trabalho são de mesma dimensão, podendo ser comparados por meio de uma métrica de distância. O método de classificação por distância Euclidiana é um procedimento de classificação supervisionada, que utiliza esta distância para associar um "vetor" a uma determinada classe. Define-se a distância Euclidiana entre dois vetores de característica A e B como sendo a raiz quadrada do somatório do quadrado das diferenças [7].

$$d(A, B) = \left( \sum_{i=1}^9 [A[i] - B[i]]^2 \right)^{1/2} \quad (1)$$

Esta métrica foi utilizada entre pares de vetores de características, buscando medir a similaridade entre o vetor da imagem sob análise e os vetores das imagens do banco de conhecimento. O valor da distância Euclidiana entre dois vetores mostra a proximidade da imagem sob análise pertencer a algum dos padrões celulares aqui estudados.

### 3. Experimentos

Na primeira fase da experimentação construiu-se o banco de conhecimento utilizando 120 imagens dos padrões celulares estudados neste trabalho. Depois de finalizado o primeiro módulo classificou-se 240 novas imagens com o objetivo de validar o processo de classificação utilizando as características extraídas por meio dos coeficientes wavelets e a métrica Euclidiana.

A primeira bateria de testes foi conduzida utilizando a transformada wavelet de Haar para extração de características. A segunda bateria de testes foi conduzida utilizando a família da transformada wavelet de Daubechies (Db2, Db4 e Db8). Para definir a classificação (laudo) foi adotado que o padrão da nova imagem iria corresponder ao padrão da imagem no banco de conhecimento que tinha a menor distância em relação à imagem sob investigação. Os resultados dos experimentos serão apresentados na seção 4.

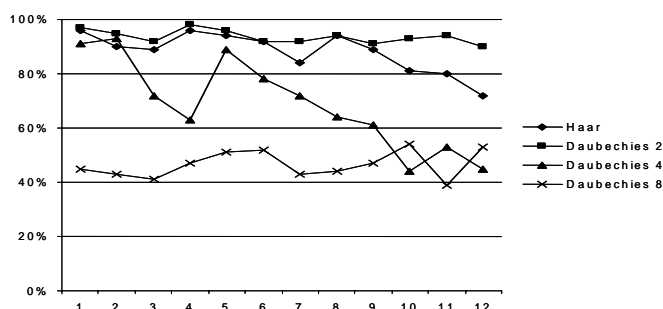
### 4. Resultados

Utilizando a transformada wavelet de Haar para o processo de classificação obteve-se em média 88,1% de classificações corretas dos padrões celulares. Utilizando a família de Daubechies obteve-se para a Db2 93,7% , Db4 68,8% e Db8 46,6% em média de classificação correta. Um resumo dos resultados é mostrado na Tabela 1.

**Tabela 1. Porcentagem de classificações corretas**

Padrões	Haar	Db2	Db4	Db8
Membrana nuclear	96%	97%	91%	45%
Homogêneo	90%	95%	93%	43%
Pontilhado pleomórfico	89%	92%	72%	41%
Pontilhado do tipo pontos isolados	96%	98%	63%	47%
Pontilhado grosso	94%	96%	89%	51%
Pontilhado grosso reticulado	92%	92%	78%	52%
Pontilhado fino	84%	92%	72%	43%
Pontilhado centromérico	94%	94%	64%	44%
Pontilhado fino denso	89%	91%	61%	47%
Nucleolar homogêneo	81%	93%	44%	54%
Nucleolar aglomerado	80%	94%	53%	39%
Nucleolar pontilhado	72%	90%	45%	53%
Média	88.1%	93.7%	68.8%	46.6%

A Figura 2 mostra o comportamento da porcentagem de acertos em cada grupo, mostrando as variações existentes em cada grupo de padrões celulares.



**Figura 2: Valores da porcentagem de classificação correta para cada tipo de padrão celular e para cada tipo de wavelet.**

Os resultados aqui apresentados refletem apenas os valores das classificações corretas encontradas experimentalmente dentre os 4 tipos de wavelets testadas.

## 5. Conclusões

Neste artigo um novo modelo de classificação de padrões de células FAN Hep-2 é apresentado. Os melhores resultados obtidos utilizando esta abordagem foi a transformada wavelet de Daubechies (Db2). A energia dos coeficientes de detalhes da transformada wavelet de Daubechies mostrou-se ferramenta útil para o processo de classificação das imagens de FAN Hep-2. Conclui-se que este método poderá definir uma possível leitura padronizada dos padrões celulares, tendo como benefício à redução dos erros de interpretação, e também podendo ser utilizada para uniformizar estudos de doenças auto-imunes que necessitam de uma leitura dos exames de forma padronizada. Este sistema poderá ser usado como auxílio na conclusão de laudos. Na prática clínica auxiliará na triagem de pacientes com suspeita de doença auto-imune, pois alcança níveis de acerto acima de 80%.

## 6. Referências

- [1] T. Pincus, J.M. Mitchell, and R.V. Burkhauser, "Substantial work disability and earnings losses in individuals less than age 65 with osteoarthritis: comparisons with rheumatoid arthritis", *J. Clin. Epidemiol.*, 1989, pp. 42:449-457.
- [2] Ferreira, A.W., S.L.M. Ávila, *Diagnóstico laboratorial das principais doenças infecciosas e auto-imunes*, E. Guanabara, Rio de Janeiro, 1996.
- [3] M.C. Arroyave, M.J. Giambrone, and K.C. Rich, "The Frequency of Antinuclear (ANA) in Children by use of Mouse Kidney (MK) and Human Epithelial Cells (Hep-2)", *Journal Allergy Clin. Immunol.*, 1988, pp. 82:741-744.
- [4] M.E. Doi, "O uso de diferentes substratos na investigação dos anticorpos antinucleares em doenças reumáticas sistêmicas", *Rev. Bras. Reumatol.*, 1995, pp. 35:16-20.
- [5] S.G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", *IEEE Trans on PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 1989, pp. 11:674-693.
- [6] I. Daubechies, *Ten lectures on wavelets*, Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- [7] Nixon, M.S., and A.S. Aguado, *Feature extraction and image processing*, Newnes, Oxford, 2002.