

TRIAGEM VIRTUAL DE IMAGENS DE IMUNO-HISTOQUÍMICA USANDO REDES NEURAIIS ARTIFICIAIS E ESPECTRO DE PADRÕES

Higor Neto Lima

Escola Politécnica de Pernambuco, Universidade de Pernambuco

higornetto@gmail.com

Wellington Pinheiro dos Santos

Núcleo de Engenharia Biomédica, Centro de Tecnologia e Geociências, Universidade Federal de Pernambuco

wellington.santos@ieee.org

Mêuser Jorge Silva Valença

Escola Politécnica de Pernambuco, Universidade de Pernambuco

mjv@ecomp.poli.br

Resumo – A importância de se organizar imagens médicas de acordo com sua natureza, aplicação e relevância tem aumentado. Além do mais, a seleção prévia de imagens médicas pode ser bastante útil para reduzir o tempo de análise dispendido por patologistas. Neste artigo é proposto um classificador de imagens para integrar um sistema de CBIR (*Content-Based Image Retrieval*, Recuperação de Imagens Baseada em Conteúdo) dedicado à seleção prévia de imagens, ou seja, à tarefa de triagem. Esse classificador é baseado na representação das imagens de entrada pelo espectro de padrões e na classificação por redes neurais. A seleção dos atributos é realizada aplicando a Análise de Componentes Principais aos vetores de atributos obtidos pela extração do espectro de padrões, enquanto a classificação de imagens é baseada no uso de redes neurais perceptron multicamadas e na composição de mapas auto-organizados e aprendizado por quantização vetorial. Esses métodos foram aplicados à seleção baseada em conteúdo de imagens de imuno-histoquímica de placenta e pulmão de neomortos. Os resultados demonstraram que esta abordagem pode atingir desempenhos de classificação razoáveis.

Palavras-chave – CBIR, k-Médias, Algoritmos de Agrupamento, Análise de Componentes Principais, Espectro de Padrões, Imuno-Histoquímica.

Abstract – The importance of organizing medical images according to their nature, application and relevance is increasing. Furthermore, a previous selection of medical images can be useful to accelerate the task of analysis by pathologists. Herein this work we propose an image classifier to integrate a CBIR (*Content-Based Image Retrieval*) selection system. This classifier is based on pattern spectra and neural networks. Feature selection is performed using pattern spectra and principal component analysis, whilst image classification is based on multilayer perceptrons and a composition of self-organizing maps and learning vector quantization. These methods were applied for content selection of immunohistochemical images of placenta and newdeads lungs. Results demonstrated that this approach can reach reasonable classification performance.

Keywords – CBIR, k-means, clustering algorithms, principal component analysis, pattern spectrum, immunohistochemistry.

1 Introdução

A velocidade de processamento dos computadores e a capacidade de armazenamento, bem como o desenvolvimento de novas mídias de armazenamento e evolução das mídias existentes, têm passado por um crescimento vertiginoso desde o começo da década de 1990 [1, 2]. Paralelamente, a evolução das tecnologias digitais tem sido cada vez mais empregada no desenvolvimento de soluções baseadas em processamento, armazenamento e análise de imagens digitais, o que, por sua vez, tem demandado mais capacidade de armazenamento e representação de imagens [3]. Aliada a essa demanda surge também a necessidade de construir métodos de busca que permitam a pesquisa de imagens baseada em conteúdo [3].

Recuperação de Imagens Baseada em Conteúdo (*Content-Based Image Retrieval*, CBIR) é uma forma de buscar imagens por meio de uma imagem-base, selecionada de acordo com o interesse do usuário. Para proporcionar a busca baseada em conteúdo, é preciso, portanto, representar as imagens no banco de dados de forma alternativa. Isso é feito usualmente por meio da construção de vetores de características que contenham os atributos de interesse para descrição e representação unívoca do conteúdo de interesse. As características mais usadas são cor, textura e forma. A recuperação de imagens baseada em conteúdo se fundamenta em três premissas básicas: extração de características visuais, indexação multidimensional, e projeto de sistemas de recuperação [4].

A representação das imagens nos bancos de dados por meio de vetores de características proporciona que o usuário possa buscar imagens na base de dados com o contexto requerido fazendo uso de uma imagem base, representada por seu vetor de características. Esse vetor é comparado aos vetores que representam as outras imagens no banco de dados e, por meio de medidas de similaridade, a pesquisa retorna as imagens de conteúdo similar [3–9]. Neste trabalho as imagens são representadas por vetores

de características obtidos da extração do espectro morfológico de padrões de cada uma das bandas da imagem colorida [10]. A dimensionalidade desses vetores é então reduzida pela aplicação da Análise de Componentes Principais (*Principal Component Analysis*, PCA).

O sistema de triagem virtual proposto neste trabalho é apresentado em duas versões: uma primeira, baseada em redes neurais MLP (*Multilayer Perceptron*), e uma segunda versão, baseada em redes SOM (*Self-Organizing Maps*) e redes LVQ (*Learning Vector Quantization*). As duas configurações são comparadas quanto ao desempenho na tarefa de triagem virtual, ou seja, quanto à taxa de acerto de pré-classificação. A escolha pelo uso de redes neurais se deve ao seu emprego em outros trabalhos envolvendo CBIR [11, 12].

Este trabalho está organizado da forma que segue: na seção ‘Materiais e Métodos’ apresentam-se a teoria básica das redes neurais artificiais do tipo perceptron multicamadas e mapas auto-organizados de Kohonen, os princípios de Morfologia Matemática e a definição do espectro de padrões, e a proposta de sistema de triagem virtual; em ‘Resultados Experimentais’ são apresentados os resultados obtidos para as duas configurações de sistema de CBIR para triagem virtual. Por fim, na seção ‘Discussão e Conclusões’ são apresentados comentários e conclusões gerais e específicas levando em conta os resultados obtidos.

2 Materiais e Métodos

2.1 Redes Perceptron de Camada Única

As redes perceptron consistem em classificadores lineares propostos por Rosenblatt que emprega neurônios MCP e uma técnica de treinamento supervisionado baseada no aprendizado de Hebb [13, 14].

Em alguns textos tal rede é considerada uma rede de duas camadas de nodos [13]. Neste trabalho, quando se referir a uma rede de M camadas, por exemplo, ficará explícito se são M camadas de nodos ou de neurônios propriamente ditos, o que exclui a camada de nodos de entrada, que não executam funções matemáticas, mas apenas distribuem as entradas para a primeira camada de neurônios.

O treinamento de uma rede perceptron de camada única é baseado no seguinte procedimento, para uma rede de m neurônios e n atributos de entrada [14]:

1. Inicializar os pesos $w_{i,j}(0)$.
2. Inicializar a taxa de aprendizado η .
3. Repetir até $\delta(t) \leq \epsilon$ ou um máximo de N_{\max} iterações, onde ϵ é o máximo erro permitido:

(a) Para cada par (\mathbf{x}, \mathbf{d}) do conjunto de treinamento Ψ , onde

$$\Psi = \{(\mathbf{x}^{(l)}, \mathbf{d}^{(l)})\}_{l=1}^L,$$

o vetor

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$$

é o vetor de atributos de entrada e

$$\mathbf{d} = (d_1, d_2, \dots, d_m)^T$$

é o vetor de saídas desejadas dos m neurônios, repetir:

- i. Calcular as saídas $y_i(t)$ dos neurônios:

$$y_i(t) = g \left(\sum_{j=1}^n w_{i,j}(t)x_j(t) \right),$$

onde $1 \leq i \leq m$.

- ii. Calcular o ajuste $\Delta w_{i,j}(t)$:

$$\delta_i(t) = d_i(t) - y_i(t),$$

$$\Delta w_{i,j}(t) = \eta \delta_i(t) x_j(t),$$

onde $1 \leq i \leq m$ e $1 \leq j \leq n$.

- iii. Atualizar os pesos:

$$w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}(t),$$

onde $1 \leq i \leq m$ e $1 \leq j \leq n$.

- (b) Calcular $\delta(t)$ da forma que segue:

$$\delta(t) = \max(\delta_1(t), \delta_2(t), \dots, \delta_m(t)).$$

As redes perceptron têm o inconveniente de apenas convergirem para casos linearmente separáveis, ou seja, casos em que existem hiperplanos que separam os conjuntos de entrada [15]. Essa limitação é resolvida com as redes perceptron multicamadas.

2.2 Redes Perceptron Multicamadas

As *redes perceptron multicamadas* foram um avanço muito importante para o fim da estagnação das pesquisas em RNAs, devido à possibilidade de, utilizando neurônios MCP e arquitetura multicamadas, resolver problemas não-linearmente separáveis [13]. Elas utilizam um algoritmo chamado *algoritmo de retropropagação*, que consiste em uma generalização do algoritmo de treinamento do perceptron de camada única através da estimação dos erros $\delta_i^{(k)}$ de uma camada k a partir dos erros $\delta_i^{(k+1)}$ da camada superior $k + 1$.

É possível provar que, para modelagem de funções, é possível modelar qualquer função contínua com uma rede perceptron de duas camadas de neurônios, enquanto qualquer função pode ser modelada usando um perceptron de três camadas de neurônios, diferente das redes de camada única, que só conseguem modelar funções contínuas e lineares [14].

O algoritmo de retropropagação, também chamado de *regra delta generalizada*, consiste no seguinte procedimento, para uma rede de $M + 1$ camadas de nodos, cada uma com N_k nodos, sendo $N_0 = n$ atributos de entrada e $N_M = m$ neurônios de saída, onde $0 \leq k \leq M$ [14]:

1. Inicializar os pesos $w_{i,j}^{(k)}(0)$.
2. Inicializar a taxa de aprendizado η .
3. Repetir até $\delta(t) \leq \epsilon$ ou um máximo de N_{\max} iterações, onde ϵ é o máximo erro permitido:

(a) Para cada par (\mathbf{x}, \mathbf{d}) do conjunto de treinamento Ψ , onde

$$\Psi = \{(\mathbf{x}^{(l)}, \mathbf{d}^{(l)})\}_{l=1}^L,$$

o vetor

$$\mathbf{x} = (x_1, x_2, \dots, x_n)^T$$

é o vetor de atributos de entrada e

$$\mathbf{d} = (d_1, d_2, \dots, d_m)^T$$

é o vetor de saídas desejadas dos $N(M) = m$ neurônios de saída, repetir:

- i. Calcular as saídas $y_i^{(k)}(t)$ dos neurônios:

$$y_i^{(k)}(t) = g \left(\sum_{j=1}^{N_{k-1}} w_{i,j}^{(k)}(t) x_j^{(k)}(t) \right),$$

onde $1 \leq i \leq N_k$ e

$$x_j^{(k)}(t) = y_j^{(k-1)}(t),$$

onde $1 \leq k \leq M$ e

$$y_j^{(0)}(t) = x_j(t),$$

para $1 \leq j \leq N_{k-1}$, $N_0 = n$ e $N_M = m$.

- ii. Calcular os erros da camada de saída:

$$\delta_i^{(M)}(t) = (d_i(t) - y_i^{(M)}(t))g'(y_i^{(M)}(t)),$$

onde $1 \leq i \leq N_M$ e $g'(\cdot)$ é a derivada de $g(\cdot)$.

- iii. Calcular os erros das outras camadas:

$$\delta_i^{(k)}(t) = g'(y_i^{(k)}(t)) \sum_{j=1}^{N_{k+1}} w_{j,i}^{(k+1)}(t) \delta_j^{(k+1)}(t),$$

onde $1 \leq i \leq N_k$ e $1 \leq k \leq M - 1$.

- iv. Calcular os ajustes $\Delta w_{i,j}^{(k)}(t)$:

$$\Delta w_{i,j}^{(k)}(t) = \eta \delta_i^{(k)}(t) x_j^{(k)}(t),$$

onde $1 \leq i \leq N_k$, $1 \leq j \leq N_{k-1}$ e $1 \leq k \leq M$.

- v. Atualizar os pesos:

$$w_{i,j}^{(k)}(t+1) = w_{i,j}^{(k)}(t) + \Delta w_{i,j}^{(k)}(t),$$

onde $1 \leq i \leq N_k$, $1 \leq j \leq N_{k-1}$ e $1 \leq k \leq M$.

(b) Calcular $\delta(t)$ da forma que segue:

$$\delta(t) = \max(\delta_1^{(M)}(t), \delta_2^{(M)}(t), \dots, \delta_{N_M}^{(M)}(t)).$$

As funções de ativação dos neurônios nas redes perceptron multicamadas devem ser contínuas, monotonicamente crescentes e diferenciáveis [16].

Existem outras versões do algoritmo de retropropagação onde os ajustes dos pesos são calculados segundo outros métodos, mas o princípio básico é o mesmo, estando a diferença apenas em questões como a rapidez da convergência e outras questões cuja natureza e interesse dependem da aplicação.

2.3 Redes SOM

Os mapas auto-organizados de Kohonen foram originalmente desenvolvidos para fazer agrupamento de dados. Também chamados de redes SOM (*Self-Organizing Maps*) [14], os mapas auto-organizados possuem uma arquitetura bastante diferente das redes neurais usuais: neles os neurônios podem ser dispostos matricialmente, tentando emular a distribuição dos neurônios biológicos no cérebro.

Nas redes SOM os neurônios podem ser dispostos de forma linear ou matricial, em todas as dimensões possíveis, sendo que a quantidade de vizinhos influencia apenas na etapa de aprendizado, estabelecendo dependências entre o ajuste dos pesos de um neurônio e os pesos de seus vizinhos, segundo uma determinada função de vizinhança, como se mostra na descrição do processo de aprendizado.

No processo de aprendizado, áreas específicas de neurônios vão sendo ativadas. Ao final do processo, a rede SOM fica dividida em áreas especialistas, responsáveis pela classificação de padrões específicos, semelhante ao cérebro, onde cada região é responsável por uma atividade específica.

Os mapas auto-organizados utilizam neurônios cuja saída é a que segue:

$$y_i = e^{-\|\mathbf{x} - \mathbf{w}_i\|^2}, \quad (1)$$

onde $\mathbf{w}_i = (w_{i,1}, w_{i,2}, \dots, w_{i,n})^T$ é o vetor de pesos do i -ésimo neurônio, para $1 \leq i \leq m$ e $1 \leq j \leq n$.

Os pesos das redes SOM podem ser ajustados pelo seguinte procedimento modificado, baseado no procedimento clássico de ajuste de redes SOM, mas tendo como critérios de parada o número máximo de iterações e a soma dos ajustes dos pesos:

1. Inicializar os pesos $w_{i,j}(0)$, onde $1 \leq i \leq m$ e $1 \leq j \leq n$, com valores aleatórios e necessariamente diferentes.
2. Entrar com os vetores posição \mathbf{r}_i , correspondentes à posição do i -ésimo neurônio na grade, onde $1 \leq i \leq m$.
3. Inicializar a taxa de aprendizado inicial η_0 , onde $0 < \eta_0 < 1$.
4. Inicializar a largura topológica σ_0 . A largura topológica é a largura da função de vizinhança gaussiana $h_{i,k}(t)$, utilizada para definir a vizinhança do neurônio vencedor a ser treinada.
5. Repetir até $\delta(t) = 0$ ou um máximo de N_{\max} iterações:

(a) Para cada padrão $\mathbf{x} = (x_1, x_2, \dots, x_n)^T$ do conjunto de entrada $\Psi = \{\mathbf{x}^{(l)}\}_{l=1}^L$, repetir:

i. Calcular as saídas $y_i(t)$ dos neurônios, onde $1 \leq i \leq m$.

ii. Calcular $y_{\max}(t)$:

$$y_{\max}(t) = \max(y_1(t), y_2(t), \dots, y_m(t)).$$

iii. Calcular o índice k do neurônio vencedor:

$$y_i = y_{\max} \Rightarrow k = i.$$

iv. Calcular o ajuste $\Delta w_{i,j}(t)$:

$$\Delta w_{i,j}(t) = \eta(t) h_{i,k}(t) (x_j(t) - w_{i,j}(t)),$$

onde $1 \leq i \leq m$ e $1 \leq j \leq n$.

v. Atualizar os pesos:

$$w_{i,j}(t+1) = w_{i,j}(t) + \Delta w_{i,j}(t),$$

onde:

$$\sigma(t) = \sigma_0 \exp\left(-\frac{t}{\tau_1}\right),$$

$$h_{i,k}(t) = \exp\left(-\frac{d_{i,k}^2}{2\sigma^2(t)}\right),$$

$$\eta(t) = \eta_0 \exp\left(-\frac{t}{\tau_2}\right),$$

onde τ_1 e τ_2 são *constantes de tempo* e:

$$d_{i,k} = \|\mathbf{r}_i - \mathbf{r}_k\|,$$

para $1 \leq i \leq m$ e $1 \leq j \leq n$.

(b) Calcular $\delta(t)$ da forma que segue:

$$\delta(t) = \sum_{i=1}^m \sum_{j=1}^n |\Delta w_{i,j}(t)|.$$

As redes SOM podem ser utilizadas também como etapa de classificação ou agrupamento de dados para uma RNA supervisionada [13, 14, 17], tendo diversas aplicações potenciais em reconhecimento de padrões na Engenharia Biomédica como, por exemplo, o auxílio ao diagnóstico da epilepsia a partir de sintomas [18].

2.4 Classificador LVQ de Kohonen

O *algoritmo LVQ*, ou algoritmo de quantização vetorial por aprendizagem (LVQ, *Learning Vector Quantization*) consiste em uma técnica supervisionada de treinamento que tanto pode ser utilizada para adicionar novos padrões a uma rede SOM já treinada quanto permite criar uma rede neural supervisionada com a mesma arquitetura de uma rede SOM [14, 19].

Dado o conjunto de treinamento $\Psi = \{\mathbf{x}^{(l)}, C^{(l)}\}_{l=1}^L$, onde $C^{(l)}$ é a classe correta *a priori*, calcula-se a saída da rede para cada elemento \mathbf{x} de Ψ pertencente à classe C e, caso o neurônio vencedor seja o i -ésimo, considerando que este neurônio está associado à classe C_i , para uma rede de n entradas e m saídas projetada para reconhecer padrões de um universo $\Omega = \{C_1, C_2, \dots, C_m\}$, onde $1 \leq i \leq m$, os pesos serão atualizados da forma que segue [14, 19]:

$$w_{i,j}(t+1) = \begin{cases} w_{i,j}(t) + \eta(t)(x_j(t) - w_{i,j}(t)), & C_i = C \\ w_{i,j}(t) - \eta(t)(x_j(t) - w_{i,j}(t)), & C_i \neq C \end{cases}, \quad (2)$$

onde $1 \leq j \leq n$ e η é a taxa de aprendizado.

A inicialização da rede LVQ é idêntica à da rede SOM. A rede LVQ pode usar como critério de parada tanto o número de iterações quanto o erro de classificação, dado que o treinamento é supervisionado.

2.5 Morfologia Matemática e Espectro de Padrões

Uma *imagem* $f : S \rightarrow K$ é definida como sendo a função que mapeia um vetor bidimensional $u \in S$ no ponto $f(u) \in K$, onde S é a *grade* ou *suporte* da imagem, u é o *pixel* da grade S ; $f(u) \in K$ é o *valor* do *pixel* $u \in S$.

De forma mais rigorosa, [20] define *pixel* como sendo o par $(u, f(u))$, ou seja, o vetor formado pela sua posição u na grade S e seu valor associado $f(u)$. Já K é o conjunto dos valores possíveis de $f(u)$. Denota-se K^S o *conjunto de todas as imagens possíveis* na grade S em K [20].

Quando a grade S é *discreta*, isto é, $S \subseteq \mathbb{Z}^2$, diz-se que $f : S \rightarrow K$ é uma *imagem discreta*. K é definido como sendo da forma $K = V^p$, onde $V \in \mathbb{R}$ é o *conjunto dos níveis de cinza* e $p \in \mathbb{N}^*$ é o *número de bandas* presentes na imagem f .

Assim, seja f uma imagem de p bandas $f : S \rightarrow V^p$, f pode ser denotada de forma equivalente por

$$f(u) = \{f_1(u), f_2(u), \dots, f_p(u)\}, \quad u \in S,$$

onde $f_j(u)$ é a j -ésima banda de $f : S \rightarrow V^p$, para $f_j : S \rightarrow V$ e $1 \leq j \leq p$.

Quando $p = 1$ diz-se que a imagem é uma imagem em *níveis de cinza* ou *monocromática*. Uma imagem é dita ser uma *imagem colorida* quando $p = 3$ e *multiespectral* quando $4 \leq p \leq 100$ [21]. Já uma imagem é dita *hiperespectral* quando $p > 100$ [21]. Para generalizar, pode-se dizer que uma imagem é multiespectral quando $p > 1$.

Uma imagem $f : S \rightarrow K$ é dita ser uma *imagem digital* quando, além de ser discreta, ela é *quantizada*, ou seja, seus *pixels* assumem apenas vetores discretos. Matematicamente, $f : S \rightarrow K$ é uma imagem digital quando $K = \{0, 1, \dots, k\}^p$ e $S \in \mathbb{Z}$, onde $p \in \mathbb{N}^*$.

Quando uma imagem digital é representada com N bits, $k = 2^N - 1$. Assim, uma imagem digital $f : S \rightarrow K$ de 8 bits tem $K = \{0, 1, \dots, 255\}^p$, enquanto que imagens digitais de 16 bits tem $K = \{0, 1, \dots, 65535\}^p$. Conseqüentemente, uma *imagem binária de banda única* $f : S \rightarrow K$ é uma imagem digital com $K = \{0, 1\}$.

Por uma questão de conveniência, é comum representar as imagens digitais como imagens discretas normalizadas, substituindo $K = \{0, 1, \dots, k\}^p$ por $\overline{K} = [0, 1]^p$. Desta forma não é preciso explicitar quantos bits são utilizados na representação dos valores possíveis dos *pixels*. Assim, uma imagem $f : S \rightarrow \{0, 1, \dots, k\}^p$ passa a ser representada por $\tilde{f} : S \rightarrow [0, 1]^p$.

É importante realçar que as imagens binárias são apenas casos particulares das imagens em níveis de cinza. Assim, uma imagem binária $f : S \rightarrow \{0, k\}^p$ passa a ser denotada utilizando a notação normalizada por $\tilde{f} : S \rightarrow \{0, 1\}^p$.

Salvo indicação em contrário, neste texto uma imagem $f : S \rightarrow K$ sempre será representada por sua versão normalizada, sendo K^S o conjunto de todas as imagens normalizadas, onde $K = [0, 1]^p$, para p bandas.

Usualmente S é uma região *retangular* de \mathbb{Z}^2 , podendo-se representar as imagens digitais em nível de cinza $f : S \rightarrow \{0, 1, \dots, k\}$ como *matrizes* [22, 23]. [23] representaria uma imagem $f : S \rightarrow V$ da seguinte forma, para S sendo uma grade $m \times n$ centrada em (r, s) :

$$f = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{(r,s)},$$

onde $a_{ij} \in V$, para $1 \leq i \leq m$ e $1 \leq j \leq n$. Isto implica dizer que, segundo esta notação:

$$f(u) = a_{ij} \Rightarrow u = (i - r, j - s), \quad u \in S.$$

Note que, neste caso, a grade S é uma região retangular de \mathbb{Z}^2 do tipo:

$$S = \{(1 - r), (2 - r), \dots, (m - r)\} \times \{(1 - s), (2 - s), \dots, (n - s)\}.$$

Esta será a notação utilizada ao longo deste texto sempre que for conveniente representar uma imagem como uma matriz.

As operações básicas entre imagens em Morfologia Matemática são a *comparação*, a *subtração*, a *intersecção* e a *união*.

É importante afirmar que as operações básicas da Morfologia Matemática só podem ser aplicadas a imagens que mantenham uma relação de ordem entre os *pixels*. Somente as imagens em níveis de cinza e as imagens binárias mantêm tal relação. No caso das imagens multiespectrais ou hiperespectrais, é necessário trabalhar banda a banda ou supor algum artifício para manter a relação de ordem. A relação de ordem é a base da Morfologia Matemática [20].

Dadas duas imagens $f_1 : S \rightarrow [0, 1]$ e $f_2 : S \rightarrow [0, 1]$, a *comparação* \leq é a operação de $[0, 1]^S \times [0, 1]^S$ em $[0, 1]^S$ definida como segue [20]:

$$(f_1 \leq f_2)(u) = \begin{cases} 1, & f_1(u) \leq f_2(u) \\ 0, & \text{c.c.} \end{cases}, \quad \forall u \in S. \quad (3)$$

Dada uma imagem $f : S \rightarrow [0, 1]$, o *negativo* de f , denotado por \bar{f} ou $\sim f$, é o operador de $[0, 1]^S$ em $[0, 1]^S$ definido como sendo [24]:

$$\bar{f}(u) = 1 - f(u), \quad \forall u \in S. \quad (4)$$

A *união* entre $f_1 : S \rightarrow [0, 1]$ e $f_2 : S \rightarrow [0, 1]$ é o operador de $[0, 1]^S \times [0, 1]^S$ em $[0, 1]^S$, denotado $f_1 \vee f_2$, definido como segue [20]:

$$(f_1 \vee f_2)(u) = \max\{f_1(u), f_2(u)\}, \quad \forall u \in S. \quad (5)$$

A *intersecção* entre $f_1 : S \rightarrow [0, 1]$ e $f_2 : S \rightarrow [0, 1]$ é o operador de $[0, 1]^S \times [0, 1]^S$ em $[0, 1]^S$, denotado $f_1 \wedge f_2$, definido como [20]:

$$(f_1 \wedge f_2)(u) = \min\{f_1(u), f_2(u)\}, \quad \forall u \in S. \quad (6)$$

Pode-se notar que as operações de união e intersecção entre imagens se assemelham às operações de mesmo nome entre conjuntos nebulosos. A semelhança entre estas operações da Morfologia Matemática e da Lógica Nebulosa cresce com a adoção da notação de imagem normalizada, o que faz com que, de certa forma, as imagens normalizadas se assemelhem a conjuntos nebulosos.

A *subtração* ou *diferença* entre $f_1 : S \rightarrow [0, 1]$ e $f_2 : S \rightarrow [0, 1]$ é operação de $[0, 1]^S \times [0, 1]^S$ em $[0, 1]^S$ dada por [24]:

$$(f_1 \sim f_2)(u) = (f_1 \wedge \bar{f}_2)(u), \quad \forall u \in S. \quad (7)$$

A partir destas operações é possível definir os *operadores básicos* da Morfologia Matemática: a *dilatação*, a *erosão*, o *gradiente morfológico*, a *abertura*, o *fechamento*, as *aberturas e fechamentos por reconstrução* e as *dilatações e erosões condicionais*.

Pode-se dizer que a Morfologia Matemática é uma teoria *construtiva*, devido ao fato de que operadores mais complexos podem ser construídos a partir dos mais simples. Destes, os operadores fundamentais são a *dilatação* e a *erosão*, originando-se todos os outros a partir de combinações destes [25].

A *dilatação* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada $\delta_g(f)$ ou $f \oplus g$ (soma de Minkowsky, no caso binário [20]), é um operador de $[0, 1]^S \times [0, 1]^S$ em $[0, 1]^S$ definido como:

$$\delta_g(f)(u) = (f \oplus g)(u) := \bigvee_{v \in S} f(v) \wedge g(u - v), \quad \forall u \in S, \quad (8)$$

onde g é denominado *elemento estruturante*. Na prática, a dilatação transforma a imagem original de forma a fazer o elemento estruturante se “encaixar” na imagem, ou seja, a imagem original f será modificada de modo a fazer as áreas semelhantes a g “crescerem”, “encaixando” g .

Visualmente, a dilatação resulta no *crescimento das áreas claras* e na *eliminação das áreas escuras menores do que o elemento estruturante*, considerando o fato de que, usualmente, os *pixels* mais claros são representados por valores mais próximos de 1, enquanto os mais escuros, por valores mais próximos de 0.

Tabela 1: Operações de dilatação, erosão e convolução

| Operação | Expressão |
|------------|---|
| Dilatação | $(f \oplus g)(u) = \bigvee_{v \in S} f(v) \wedge g(u - v)$ |
| Erosão | $(f \ominus g)(u) = \bigwedge_{v \in S} f(v) \vee \bar{g}(v - u)$ |
| Convolução | $(f * g)(u) = \sum_{v \in S} f(v)g(u - v)$ |

A erosão de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada $\epsilon_g(f)$ ou $f \ominus g$ (subtração de Minkowsky, no caso binário [20]), é um operador de $[0, 1]^S \times [0, 1]^S$ em $[0, 1]^S$ definido como:

$$\epsilon_g(f)(u) = (f \ominus g)(u) := \bigwedge_{v \in S} f(v) \vee \bar{g}(v - u), \quad \forall u \in S. \quad (9)$$

Na prática, a erosão transforma a imagem original f de modo a fazer as áreas semelhantes a g “diminuírem”, “encaixando” \bar{g} , conforme o exemplo anterior.

Visualmente, a erosão resulta no *crescimento das áreas escuras* e na *eliminação das áreas claras menores do que o elemento estruturante*, considerando o fato de que, usualmente, os *pixels* mais claros são representados por valores mais próximos de 1, enquanto os mais escuros, mais próximos de 0.

Em [24] a dilatação e a erosão são mostradas esquematicamente por meio de exemplos e ilustrações, que mostram o elemento estruturante percorrendo uma imagem binária. É uma abordagem bastante ilustrativa, que dá ao leitor uma idéia de como são implementadas de forma algorítmica a dilatação e a erosão.

A Tabela 1 mostra as expressões das operações de *dilatação*, *erosão* e *convolução*, podendo-se perceber a semelhança entre estas três operações.

Quanto às suas propriedades, a *dilatação* é uma *operação extensiva* devido ao fato de $\delta_g(f)(u) \geq f(u)$, $\forall u \in S$, enquanto que a *erosão* é uma operação *anti-extensiva*, pois $\epsilon_g(f)(u) \leq f(u)$, $\forall u \in S$ [10].

A *anti-dilatação* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada por $\delta_g^a(f)$, é dada por:

$$\delta_g^a(f)(u) := \overline{\delta_g(f)}(u), \quad \forall u \in S. \quad (10)$$

A *anti-erosão* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada por $\epsilon_g^a(f)$, é dada por:

$$\epsilon_g^a(f)(u) := \overline{\epsilon_g(f)}(u), \quad \forall u \in S. \quad (11)$$

A *n-dilatação* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada por $\delta_g^n(f)$, é dada por, para $n > 1$:

$$\delta_g^n(f)(u) := \underbrace{\delta_g \delta_g \cdots \delta_g}_{n} (f)(u), \quad \forall u \in S. \quad (12)$$

De forma semelhante, a *n-erosão* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada por $\epsilon_g^n(f)$, é dada por, para $n > 1$:

$$\epsilon_g^n(f)(u) := \underbrace{\epsilon_g \epsilon_g \cdots \epsilon_g}_{n} (f)(u), \quad \forall u \in S. \quad (13)$$

O elemento estruturante *quadrado* 3×3 é dado por:

$$b = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}_{(2,2)}.$$

O elemento estruturante *cruz* 3×3 é dado por:

$$b = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix}_{(2,2)}.$$

Em [24] é demonstrado que qualquer elemento estruturante de dimensões maiores que 3×3 pode ser decomposto em vários 3×3 . Isto quer dizer que uma dilatação por um elemento estruturante h de dimensões $m \times n$, tal que $m > 3$ e $n > 3$, pode ser executada por meio de um certo número de dilatações por b_k de dimensões 3×3 , onde cada b_k é um elemento estruturante que constitui h [24].

Uma das aplicações mais importantes da Morfologia Matemática é a implementação de *filtros morfológicos*, os quais consistem em filtros não lineares baseados em operações morfológicas. Os filtros morfológicos mais comuns são a *abertura*, o *fechamento*, a *n-abertura*, o *n-fechamento* e os *filtros alternados*.

Segundo [20], uma transformação ξ é um *filtro morfológico* se for ao mesmo tempo *idempotente* e *crescente*.

A transformação ξ é *crecente* se, dadas duas imagens $f : S \rightarrow K$ e $g : S \rightarrow K$, a seguinte condição é satisfeita [20]:

$$f(u) \leq g(u) \Rightarrow \xi(f)(u) \leq \xi(g)(u), \quad \forall u \in S. \quad (14)$$

Por sua vez, a transformação ξ é *idempotente* se [20]:

$$\xi\xi(f)(u) = \xi(f)(u), \quad \forall u \in S, \quad (15)$$

ou seja, se ξ é idempotente, a imagem f não é alterada por aplicações sucessivas da mesma transformação ξ .

O filtro morfológico de *abertura* é utilizado para “eliminar” regiões claras na imagem que não possam estar contidas no elemento estruturante. Assim, a abertura realiza na imagem uma espécie de “peneiramento” dos pontos claros. Como resultado, fazendo uma analogia entre uma imagem em níveis de cinza e um relevo topográfico, os picos (pontos de máximo ou regiões mais claras) inferiores em tamanho ao elemento estruturante são “eliminados”.

Matematicamente, a *abertura* de $f : S \rightarrow [0, 1]$ pelo elemento estruturante $g : S \rightarrow [0, 1]$, denotada $\gamma_g(f)$, é a transformação de $[0, 1]^S$ em $[0, 1]^S$ definida por [20]:

$$\gamma_g(f) := \delta_g \epsilon_g(f). \quad (16)$$

A abertura $\gamma_g(f)$ é uma transformação *anti-extensiva*, pois:

$$\gamma_g(f)(u) \leq f(u), \quad \forall u \in S. \quad (17)$$

O filtro morfológico de *fechamento* é utilizado para “eliminar” regiões escuras na imagem que não possam estar contidas no elemento estruturante. Assim, o fechamento realiza na imagem uma espécie de “peneiramento” dos pontos escuros. Como resultado, novamente utilizando uma analogia entre uma imagem em níveis de cinza e um relevo topográfico, os vales (pontos de mínimo ou regiões mais escuras) inferiores em tamanho ao elemento estruturante são “eliminados”.

Matematicamente, o *fechamento* de $f : S \rightarrow [0, 1]$ pelo elemento estruturante $g : S \rightarrow [0, 1]$, denotado $\phi_g(f)$, é a transformação de $[0, 1]^S$ em $[0, 1]^S$ definida da forma que segue [20]:

$$\phi_g(f) := \epsilon_g \delta_g(f). \quad (18)$$

O fechamento $\phi_g(f)$ é uma transformação *extensiva*, pois:

$$f(u) \leq \phi_g(f)(u), \quad \forall u \in S. \quad (19)$$

A *n-abertura* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotada $\gamma_g^n(f)$, é definida da forma que segue [20]:

$$\gamma_g^n(f) := \delta_g^n \epsilon_g^n(f). \quad (20)$$

O *n-fechamento* de $f : S \rightarrow [0, 1]$ por $g : S \rightarrow [0, 1]$, denotado $\phi_g^n(f)$, é definido como segue [20]:

$$\phi_g^n(f) := \epsilon_g^n \delta_g^n(f). \quad (21)$$

Dentre as ferramentas da Morfologia Matemática para a descrição de tamanho e forma em análise de imagens, as *granulometrias* constituem algumas das mais utilizadas. A idéia básica de uma granulometria é realizar diversas *crivagens* (ou *peneiramentos*) seguidas de medições quantitativas dos resíduos, tais como medição de área ou perímetro. Ou seja, realiza-se um *peneiramento* da imagem original utilizando uma certa *peneira* e depois medem-se os *grãos* que passaram; em seguida, utiliza-se uma *peneira* com *furos menores* e novamente se medem os resíduos, e assim por diante. No caso, o *tamanho dos furos da peneira* é determinado pela dimensão do elemento estruturante utilizado na transformação granulométrica.

A família de transformações $\{\psi_j\}$, parametrizadas com um parâmetro $j \geq 0$, onde

$$\psi_0(f)(u) = f(u), \quad \forall u \in S, \quad f : S \rightarrow [0, 1],$$

é chamada de uma *granulometria* se constituir um *critério de forma* [10], ou seja:

$$\psi_j(f)(u) \leq f(u), \quad (22)$$

$$f(u) \leq g(u) \Rightarrow \psi_j(f)(u) \leq \psi_j(g)(u), \quad (23)$$

$$\psi_j[\psi_k(f)](u) = \psi_k[\psi_j(f)](u) = \psi_{\max(j,k)}(f)(u), \quad \forall j, k \geq 0, \quad (24)$$

$\forall u \in S$, onde $g : S \rightarrow [0, 1]$.

A última condição expressa a idéia intuitiva de que dois *peneiramentos* seguidos são equivalentes a um só *peneiramento* com a *peneira de buracos menores* ($\max(j, k)$), já que o índice j da transformação ψ_j expressa o rigor do *peneiramento*: quanto maior o j , mais rigoroso o *peneiramento*, ou seja, menores os *grãos* passantes.

É possível demonstrar que o conjunto de transformações morfológicas que constitui um *critério de forma* é o conjunto das *j-aberturas* por um elemento estruturante *disco digital* [10]. Os discos digitais mais elementares são o quadrado 3×3 e a cruz 3×3 [10]. Assim, $\{\gamma_g^j\}$ é uma *granulometria*, onde $j \geq 0$ e g é um disco digital.

Logo, uma granulometria passa a ser expressa da forma $\{\psi_j = \gamma_g^j\}$, $r \geq 0$, onde g é um disco digital [10].

Analogamente, define-se uma *anti-granulometria* como sendo o conjunto dos j -fechamentos $\{\phi_g^j\}$, onde $j \geq 0$ e g é um disco digital [10].

Seja uma imagem $f : S \rightarrow [0, 1]$. A função $V : \mathbb{Z}_+ \rightarrow \mathbb{R}$ é definida como sendo [10]:

$$V(k) = \sum_{u \in S} \gamma_g^k(f)(u), \quad (25)$$

onde g é a cruz ou o quadrado 3×3 . Assim, a função $\Xi : \mathbb{Z}_+ \rightarrow [0, 1]$ é definida da forma que segue:

$$\Xi[k] = 1 - \frac{V(k)}{V(0)}, \quad k \geq 0. \quad (26)$$

Considerando a imagem f um *conjunto aleatório* e observando que Ξ é uma função monotônica, crescente e limitada em $[0, 1]$, pois $V(k+1) < V(k)$, $\forall k \geq 0$, podemos afirmar que Ξ é uma *função de distribuição acumulada discreta* associada à imagem f [10]. Assim, pode-se definir uma *função densidade discreta* $\xi : \mathbb{Z}_+ \rightarrow \mathbb{R}_+$ por meio da diferença:

$$\xi[k] = \Xi[k+1] - \Xi[k], \quad k \geq 0. \quad (27)$$

A função densidade definida na equação 27 é denominada *espectro de padrões* [10]. O espectro de padrões, ou *espectro morfológico*, é uma espécie de *histograma de tamanho e forma*, sendo *único* para uma imagem específica f e um determinado elemento estruturante g , desde que f seja *binária* [10].

Uma vez que as transformações granulométricas têm a propriedade de “espalhar” informações de tamanho e forma nas imagens resíduo, muitas outras representações diferentes do histograma de padrões dos objetos de interesse podem ser geradas, desde que uma quantidade adequada de transformações seja utilizada. Como é de se esperar, o número de transformações (ou *peneiramentos*) necessário, ou seja, o maior valor de k , varia de acordo com a aplicação, pois está diretamente associado à dimensão do vetor de atributos a ser utilizado na aplicação de reconhecimento de padrões.

2.6 Triagem virtual

A Triagem Virtual consiste em agrupar imagens médicas semelhantes para facilitar a sua distribuição para os especialistas responsáveis pela análise de imagens e, por conseguinte, pelo diagnóstico. Neste trabalho foi estudada uma aplicação baseada em imagens de imuno-histoquímica da placenta e do pulmão do tipo CD68 [26, 27]. Essas imagens são utilizadas para se poder auxiliar o patologista no seu diagnóstico fazendo comparação com outros pacientes e seus respectivos diagnósticos. Fazendo a comparação visual ou com o auxílio de algum programa o patologista pode analisar o caso atual de maneira a reduzir a probabilidade de erro.

A busca de imagens em uma base de imagens médicas pode ser prejudicada por fatores como os que seguem:

1. As imagens foram agrupadas de maneira subjetiva: influência da subjetividade de cada patologista na análise pode gerar inconsistências no agrupamento;
2. A recuperação das imagens pode retornar um número insuficiente de imagens ou um conjunto de imagens com pouca relevância com o problema;
3. O sistema pode demorar muito para recuperar imagens relevantes: mesmo que o sistema recupere uma quantidade grande de imagens relevantes, a quantidade de tempo gasta com essa recuperação também é importante para o usuário.

Neste trabalho as imagens são representadas por meio de vetores de atributos construídos usando uma concatenação dos espectros de padrões das bandas R, G e B que compõem as imagens coloridas; em seguida são feitos estudos usando PCA para redução da dimensionalidade do vetor de atributos, objetivando a redução do custo computacional na tarefa de classificação. Assim, a metodologia de busca da solução do problema de triagem virtual estudado neste trabalho foi dividida em cinco etapas: (1) pré-processamento, (2) segmentação das imagens, (3) construção dos espectros de padrões, (4) análise PCA sobre os espectros e (5) uso de SOM-LVQ e MLP para classificar os vetores de características. Essas etapas são ilustradas no esquema da Figura 1. Na última fase foram utilizados conjuntos de treino e teste para poder avaliar quantitativamente a taxa de acertos na triagem das imagens.

A segmentação das imagens R, G e B foi feita com uma implementação de k -médias com 4 classes. As imagens utilizadas foram analisadas e na média existiam de 3 a 6 classes com o maior número de *pixels*. Dentre essas classes, apenas 4 seriam suficientes para representar as imagens. Com isso, foram escolhidas 3 classes para representar os objetos presentes na imagem, enquanto uma quarta classe representa fundo da imagem. Ao fim da segmentação a quarta classe é removida, restando apenas três classes.

Foram calculados os espectros de padrões para as imagens R, G e B usando 25 iterações para compor cada espectro, ficando assim cada banda espectral com um espectro de 25 atributos. Em seguida esses espectros foram concatenados em um vetor único, obtendo-se ao fim um vetor de 75 dimensões.

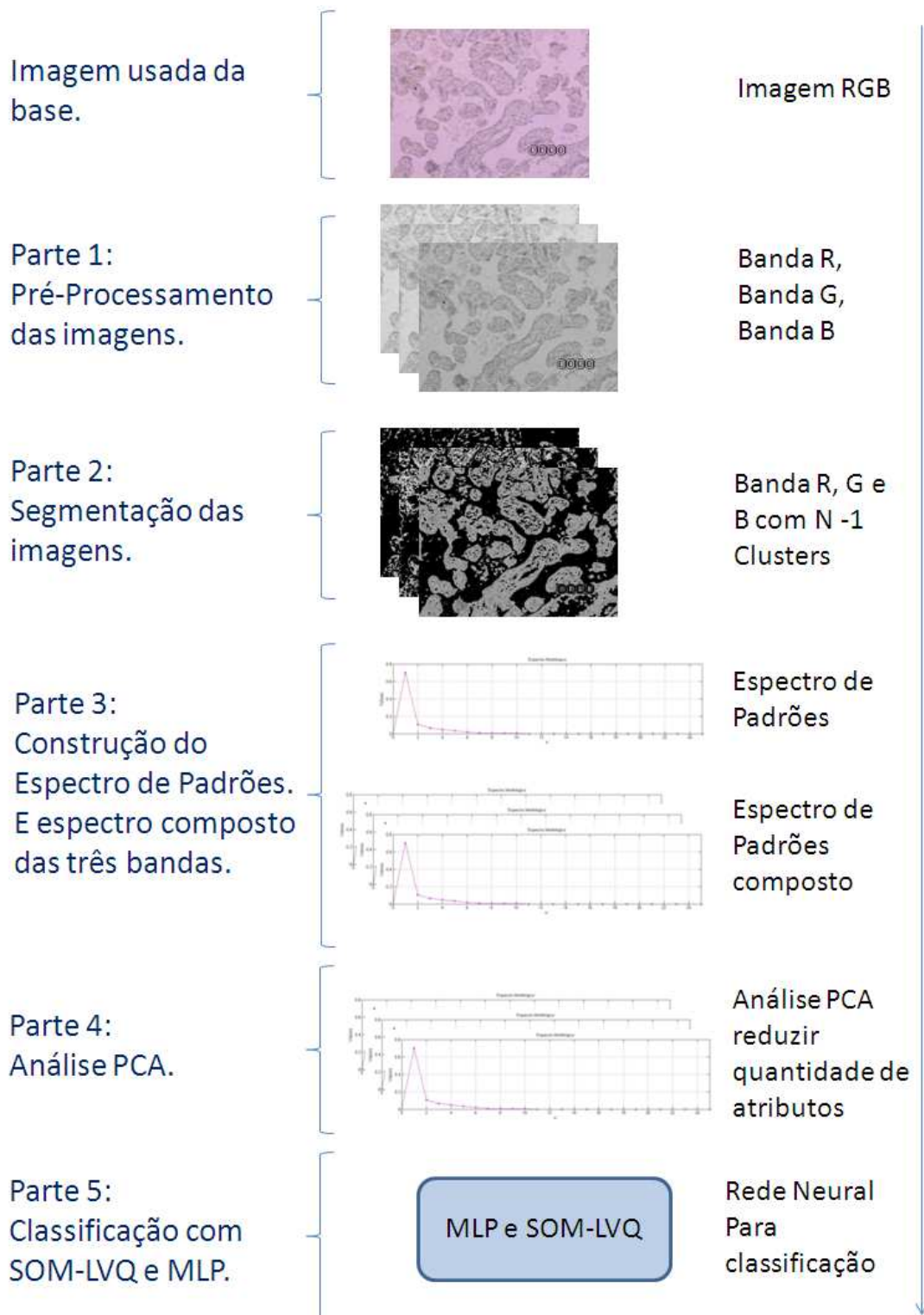


Figura 1: Metodologia empregada: etapas da classificação para triagem virtual

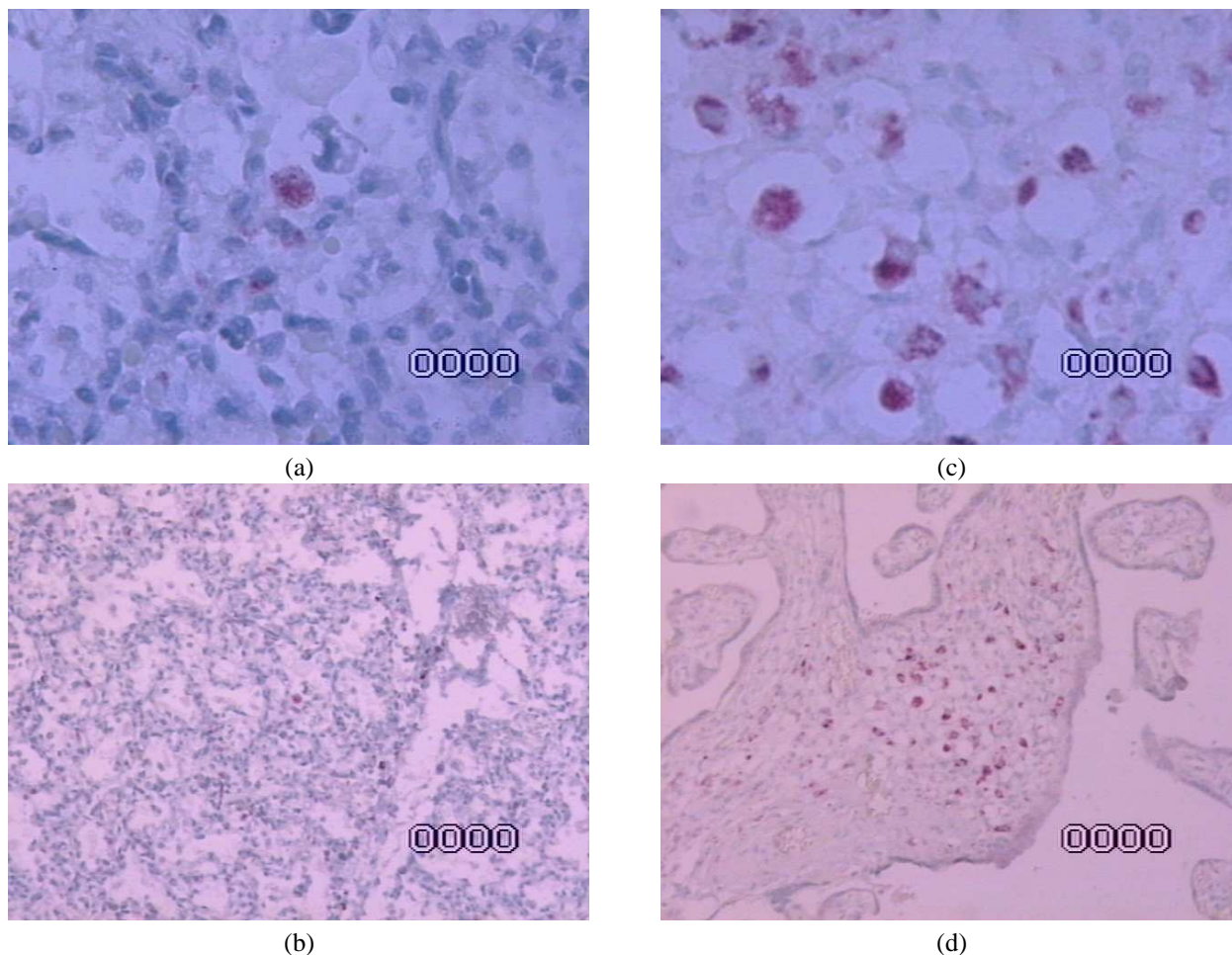


Figura 2: Tipos de imagens de imuno-histoquímica CD68: (a) NP com mais *zoom*; (b) NP com menos *zoom*; (c) P com mais *zoom*; e (d) P com menos *zoom*.

A análise PCA foi o passo que teve como entrada a matriz unificada gerada, que representava as imagens dos dois tipos. Essa matriz foi processada para obter a menor representação, i.e. com a menor quantidade de atributos, mas mantendo uma boa representação dos dados.

Foi usada uma rede SOM-LVQ e uma rede MLP para classificar os dados usando a validação cruzada como condição de parada da rede. As duas redes foram testadas usando a matriz resultante da análise PCA e a matriz sem o uso da análise PCA. A classificação foi dividida em duas etapas, como descrito adiante.

Primeiro foi feita uma análise para se descobrir a melhor configuração de rede para o problema. Para isso foi estabelecida uma bateria de testes que fez os treinamentos da rede variando a quantidade de neurônios na camada escondida. Foram utilizados de 10 a 30 neurônios com 20 repetições para cada rede. Em cada passo foi guardada a melhor rede, baseando-se na taxa de acertos e na média de acertos de todas as redes. O segundo passo foi repetir os testes com a melhor rede e obter a média, desvio padrão e mediana da taxa de acertos.

A matriz de entrada foi dividida em 19% para teste e 81% para treino e os valores foram embaralhados no início dos treinos. Com isso, o resultado obtido foi uma comparação da abordagem usando o PCA, para reduzir a complexidade, mas com o intuito de manter uma boa representação das imagens com uma abordagem de referência usando todos os parâmetros do vetor de entrada.

Neste trabalho foram utilizadas imagens de imuno-histoquímica CD68 da placenta e do pulmão [26, 27]. As imagens foram diferenciadas pela nomenclatura P (pulmão) e NP (placenta). As imagens foram adquiridas usando dois tipos de *zoom*. A Figura 2 ilustra quatro de todas as imagens utilizadas. A base de dados tem um total de 113 imagens, sendo 53 imagens do tipo NP e 60 imagens do tipo P.

3 Resultados Experimentais

As imagens na aplicação de CBIR de Triagem Virtual foram representadas usando vetores multidimensionais baseados nos seguintes passos, e com os resultados que seguem:

1. Cada imagem foi separada em bandas R, G e B;
2. Cada banda foi segmentada, resultando uma imagem com quatro clusters, sendo o maior retirado;

- Foi construído o espectro de padrões da cada banda resultando em um vetor de 25 dimensões. Depois de serem unificados, resultaram em um vetor de 75 dimensões. As classes de imagens NP e P ficaram com matrizes 53 por 75 e 60 por 75, respectivamente. Por fim, foi gerada a matriz de características para representar a base de dimensão 113 por 75 pela união dessas duas matrizes;
- A análise PCA reduziu a complexidade da base para uma matriz 113 por 50. Obteve-se assim uma redução de 33,33% na quantidade de atributos da matriz. Ou seja, uma redução de cerca de 1/3.

Os testes de agrupamento das imagens foram feitos em duas etapas. Para implementar o método proposto para triagem virtual foi usado o ambiente Matlab na versão R2007b como plataforma para construção, teste e treinamento da metodologia proposta. Foram utilizados 22 padrões para teste e 91 padrões para treinamento. Desses valores de teste, 11 eram do tipo NP e 11 eram do tipo P. Dos 91 padrões para treinamento, 42 eram do tipo NP e 49 eram do tipo P. Todos os treinamentos usaram validação cruzada com parte dos dados de treinamento.

Após a primeira etapa de classificação foram geradas estatísticas para embasar a seleção da melhor rede. Os resultados são mostrados na Tabela 2, com a rede utilizada, a quantidade de atributos que foram usados como entrada da rede (m), a taxa de acerto da melhor rede (η_{OTM}), a média ($\bar{\eta}$), o desvio padrão ($\Delta\eta$), a mediana (η_{MED}) e a quantidade de neurônios da melhor rede (n_{OTM}).

Tabela 2: Resultados da primeira etapa da classificação, usando PCA ($m = 50$) e sem uso do PCA ($m = 75$)

| Rede | m | η_{OTM} (%) | $\bar{\eta} \pm \Delta\eta$ (%) | η_{MED} (%) | n_{OTM} |
|---------|-----|------------------|---------------------------------|------------------|-----------|
| MLP | 50 | 95,23 | $57,05 \pm 0,48$ | 68,18 | 22 |
| | 75 | 90,47 | $53,83 \pm 0,47$ | 50,00 | 19 |
| SOM LVQ | 50 | 66,66 | $50,49 \pm 0,23$ | 40,91 | 10 |
| | 75 | 80,95 | $47,14 \pm 0,15$ | 50,00 | 17 |

Nas Tabelas 2 e 3 são apresentados os resultados para MLP e SOM-LVQ, considerando quantidades de atributos 50 e 75, com e sem o uso de PCA, respectivamente. Para as redes MLP usando padrões obtidos com o uso de PCA ($m = 50$) foi observada uma média de acertos de 57,05% e um desvio padrão de 0,48%; no entanto, foi observada uma mediana distante da média (mediana de 68,18%), indicando que houve resultados díspares em relação à média. Isso fica evidente quando se observa a taxa de acerto da melhor rede, de 95,23%. Neste caso, a rede com 22 neurônios na camada escondida foi escolhida como a melhor rede, baseando-se na sua alta taxa de acertos. No caso das redes MLP sem usar padrões obtidos com PCA ($m = 75$) foi observada uma média de acertos de 53,83% e um desvio padrão de 0,47%; porém não foi observada uma mediana tão distante da média (mediana de 50,00%), indicando que os resultados estavam mais simétricos. Neste caso, a rede com 19 neurônios na camada escondida foi escolhida como a melhor rede, baseando-se na sua alta taxa de acertos de 90,47%.

Para as redes SOM-LVQ com o uso de PCA ($m = 50$) foi observada uma redução na média de acertos das redes em comparação com as redes MLP. Com uma média de acertos de 50,49% e um desvio padrão de 0,23%. Como nas redes anteriores, foi observada uma mediana distante da média (mediana de 40,91%), indicando que houve resultados díspares em relação à média. Contudo, o baixo valor do desvio padrão indica que os valores das taxas de acertos para as redes não oscilaram muito em relação à média. Assim, a rede com 10 neurônios na camada escondida foi escolhida como a melhor rede, com uma taxa de acerto de 66,66%. Para as redes SOM-LVQ, sem usar padrões obtidos com PCA ($m = 75$), observou-se também uma redução da efetividade de acertos da rede em comparação com as redes MLP, com uma média de acertos de 47,14% e um desvio padrão de 0,15, com uma mediana não muito distante da média (mediana de 50,00%) indicando que não houve resultados díspares em relação à média, embora possa ser notado com clareza que a taxa de acerto da melhor rede foi bastante alta: 81%. Neste caso, a rede com 17 neurônios na camada escondida foi escolhida como a melhor rede, baseando-se na sua alta taxa de acertos. Depois de selecionar as melhores redes, na segunda etapa foram obtidos os resultados da Tabela 3.

Tabela 3: Resultados da segunda etapa da classificação, usando PCA ($m = 50$) e sem uso do PCA ($m = 75$)

| Rede | m | $\bar{\eta} \pm \Delta\eta$ (%) | η_{MED} (%) | η_{OTM} (%) |
|---------|-----|---------------------------------|------------------|------------------|
| MLP | 50 | $60,95 \pm 0,47$ | 70,45 | 80,95 |
| | 75 | $55,71 \pm 0,49$ | 68,18 | 80,95 |
| SOM LVQ | 50 | $52,85 \pm 0,21$ | 36,36 | 61,90 |
| | 75 | $49,76 \pm 0,15$ | 45,45 | 71,42 |

Para as redes MLP usando padrões obtidos com PCA ($m = 50$), foi observada uma média de acertos de 60,95%, um desvio padrão de 0,47%, e uma mediana de 70,45%, indicando uma certa distância entre as taxas de acertos das redes. Isso fica mais claro quando se observa a taxa de acerto da melhor rede, de 80,95%. No caso das redes MLP sem usar padrões obtidos com PCA ($m = 75$), obteve-se uma média de acertos de 55,71%, desvio padrão de 0,49% e mediana de 68,18%. A melhor rede teve uma taxa de acerto de 80,95%. Foi obtido, assim, um resultado parcialmente similar àquele obtido para as redes com PCA, porém com uma média de acertos inferior.

Para as redes com SOM-LVQ usando padrões com PCA ($m = 50$), novamente foi observada uma redução na média de acertos das redes em comparação com as redes MLP. Observou-se uma taxa de acertos de 52,85%, um desvio padrão de 0,21%, e uma mediana de 36,36%. O baixo desvio padrão e o valor da mediana sugerem uma razoável assimetria entre os dados. Contudo

a taxa de acerto da melhor rede não ficou tão distante da média: 61,90%. No caso das redes SOM-LVQ sem usar padrões obtidos com PCA ($m = 75$), obteve-se uma média de acertos de 49,76%, uma média de 0,15%, e mediana de 45,45%. Porém, mesmo com o valor relativamente baixo de desvio padrão e a proximidade entre mediana e média, obteve-se uma melhor rede com uma taxa de acerto de 71,42%, ficando acima da melhor rede SOM-LVQ com PCA.

4 Discussão e Conclusões

O uso de redes MLP no método de triagem virtual proposto neste trabalho resultou numa taxa de acerto de 81%, o que pode ser considerado razoável, uma vez que a triagem virtual desempenha o papel de pré-classificação. Contudo, para validar a proposta, seria necessário realizar uma triagem usando especialistas humanos, especificamente patologistas. Este trabalho busca na verdade ser uma prova de conceito, mostrando que é possível aplicar a ideia de CBIR para desenvolver um sistema de triagem virtual para otimizar o trabalho de análise de imagens realizado por especialistas humanos, usando o espectro de padrões para representar as imagens, sendo este o maior diferencial deste trabalho em relação a outras aplicações de CBIR.

Os resultados obtidos indicam que o uso de PCA para reduzir a quantidade de atributos suficiente para representar os dados reduziu o custo computacional, dado que reduziu a dimensionalidade, e aumentou a taxa de acerto da classificação. Nas redes MLP houve uma melhora de aproximadamente 9% da taxa de acerto com o uso de PCA. Já nas redes SOM-LVQ houve uma melhora de 6% em relação à taxa de acertos com o uso de PCA.

As redes MLP se mostraram mais adequadas ao problema, com alta taxa de acerto nas melhores redes: 81%. As redes SOM-LVQ obtiveram resultados não tão razoáveis nas suas classificações. As redes obtiveram uma melhor rede com uma taxa de 71,4% sem usar PCA superior a melhor rede com o uso de PCA, indicando que, nesse caso, as redes SOM-LVQ necessitam de mais informações para poder ter uma melhor generalização e por conseguinte se adequar ao problema em questão. A comparação dos resultados da triagem usando MLP e SOM mostrou que as redes MLP se adequaram melhor ao problema, com uma maior taxa de acerto.

Uma das dificuldades encontradas foi o alto custo computacional aparente necessário para obtenção dos espectros de padrões das imagens, já que era calculado o espectro para cada banda de uma imagem e isso era repetido para toda a base. Depois de gerado os vetores, o treinamento para se obter as melhores redes exigia um tempo de processamento também razoável. Como trabalho futuro podem ser investigadas arquiteturas paralelas para otimização do tempo dispendido no cálculo dos espectros de padrões, buscando paralelizar ao máximo a execução das granulometrias.

Referências

- [1] G. Gagaudakis and P. L. Rosin. “Shape measures for image retrieval”. *Pattern Recognition Letters*, vol. 24, no. 15, pp. 2711–2721, 2003.
- [2] G. Gagaudakis and P. L. Rosin. “Incorporating shape into histograms for CBIR”. *Pattern Recognition*, vol. 35, no. 1, pp. 81–91, 2002.
- [3] C. A. B. Castañón. “Recuperação de imagens por conteúdo através de análise multiresolução de Wavelets”. Master’s thesis, Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo, São Paulo, Brasil, 2003.
- [4] T. S. H. Y. Rui and S. F. Chang. “Image Retrieval: Current Techniques, Promising Directions, and Open Issues”. *Journal of Visual Communication and Image Representation*, vol. 10, no. 1, pp. 39–62, 1999.
- [5] A. Zhang, R. S. Aygün and Y. Song. “Feature-based retrieval in visual database systems”. In *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, edited by D. Feng, W. C. Siu and H. J. Zhang, New York, 2003. Springer-Verlag.
- [6] H. K. Huang. “PACS, image management and imaging informatics”. In *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, edited by D. Feng, W. C. Siu and H. J. Zhang, New York, 2003. Springer-Verlag.
- [7] T. W. Cai, D. D. Feng and R. Fulton. “Content-based retrieval for medical data”. In *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, edited by D. Feng, W. C. Siu and H. J. Zhang, New York, 2003. Springer-Verlag.
- [8] F. Long, H. Zhang and D. D. Feng. “Fundamentals of content-based image retrieval”. In *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, edited by D. Feng, W. C. Siu and H. J. Zhang, New York, 2003. Springer-Verlag.
- [9] J. S. jin. “Indexing and retrieving high dimensional visual features”. In *Multimedia Information Retrieval and Management: Technological Fundamentals and Applications*, edited by D. Feng, W. C. Siu and H. J. Zhang, New York, 2003. Springer-Verlag.
- [10] U. M. Braga Neto. “Reconstrução Volumétrica e Análise de Imagens Tridimensionais por Morfologia Matemática”. Master’s thesis, Faculdade de Engenharia Elétrica da Universidade Estadual de Campinas, Campinas, Brasil, 1994.

- [11] A. Antani, L. R. Long, G. R. Thoma and R. J. Stanley. “Vertebra shape classification using MLP for content-based image retrieval”. In *Proceedings of the International Joint Conference on Neural Networks*, pp. 160–165. IEEE, 2003.
- [12] K. K. Seo. “An application of one-class support vector machines in content-based image retrieval”. *Expert Systems with Applications*, vol. 33, no. 2, pp. 491–498, 2007.
- [13] Z. L. Kovács. *Redes Neurais Artificiais: Fundamentos e Aplicações*. Acadêmica São Paulo, São Paulo, 1996.
- [14] A. P. Braga, A. C. P. L. F. Carvalho and T. B. Ludermir. *Redes Neurais Artificiais: Teoria e Aplicações*. LTC, São Paulo, 2000.
- [15] J. Sklansky and G. N. Wassel. *Pattern Classifiers and Trainable Machines*. Springer-Verlag, first edition, 1981.
- [16] S. Arik. “An Analysis of Exponential Stability of Delayed Neural Networks with Time Varying Delays”. *Neural Networks*, no. 17, pp. 1027–1031, 2004.
- [17] F. M. Ham and I. Kostanic. *Principles of Neurocomputing for Science and Engineering*. McGraw-Hill, 2001.
- [18] F. A. Sala, F. M. Azevedo and F. I. M. Argoud. “Mapas de Kohonen na detecção de eventos epileptogênicos”. In *III Congresso Latino-Americano de Engenharia Biomédica*, pp. 533–536, João Pessoa, Brasil, 2004. Sociedade Brasileira de Engenharia Biomédica.
- [19] S. Haykin. *Redes Neurais: Princípios e Prática*. Bookman, Porto Alegre, 2001.
- [20] A. L. B. Candeias. “Aplicação da Morfologia Matemática à Análise de Imagens de Sensoriamento Remoto”. Ph.D. thesis, Departamento de Computação Aplicada do Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Brasil, 1997.
- [21] D. Landgrebe. “Hyperspectral Image Analysis”. *IEEE Signal Processing Magazine*, Jan 2002.
- [22] R. Klette and P. Zamperoni. *Handbook of Image Processing Operators*. John Wiley & Sons, first edition, 1995.
- [23] E. R. Dougherty and C. R. Giardina. *Matrix Structured Image Processing*. Prentice Hall, 1987.
- [24] G. J. F. Banon and J. Barrera. *Bases da Morfologia Matemática para Análise de Imagens Binárias*. UFPE-DI, 1994.
- [25] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag, New York, 2004.
- [26] W. P. Santos. “Análise de Imagens Digitais em Patologia utilizando Morfologia Matemática e Lógica Nebulosa”. Master’s thesis, Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Pernambuco, Recife, Brasil, 2003.
- [27] W. P. Santos, A. L. B. Candeias, A. F. S. Bezerra and A. F. D. Silva. “Auxílio ao Diagnóstico da Doença da Membrana Hialina utilizando Análise de Imagens por Lógica Nebulosa”. In *III Congresso Latino-Americano de Engenharia Biomédica*, pp. 1243–1246, João Pessoa, Brasil, 2004. Sociedade Brasileira de Engenharia Biomédica.