

# AVALIAÇÃO DO MÉTODO DIALÉTICO NA QUANTIZAÇÃO DE IMAGENS MULTIESPECTRAIS

Wellington Pinheiro dos Santos

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

wellington.santos@ieee.org

Francisco Marcos de Assis

Departamento de Engenharia Elétrica, Universidade Federal de Campina Grande

fmarcos@dee.ufcg.edu.br

**Resumo** – A classificação não supervisionada tem um papel muito importante na análise de imagens multiespectrais, dada a sua capacidade para auxiliar a extração de conhecimento *a priori* de imagens. Algoritmos como k-médias e *fuzzy* c-médias tem sido muito utilizados nessa tarefa. A Inteligência Computacional tem-se mostrado como um importante campo para auxiliar na construção de classificadores otimizados, tanto quanto à qualidade do agrupamento de classes, quanto à avaliação da qualidade da quantização vetorial. Diversos trabalhos têm mostrado que a Filosofia, em especial o Método Dialético, tem servido como importante inspiração para a construção de novos métodos computacionais. Este trabalho apresenta uma avaliação de quatro métodos baseados na Dialética: o Classificador Dialético Objetivo e o Método Dialético de Otimização adaptado à construção de uma versão do k-médias otimizada segundo índices de qualidade de agrupamento, cada um desses em uma versão canônica e em outra versão obtida pela aplicação do Princípio da Máxima Entropia. Esses métodos foram comparados aos métodos k-médias, *fuzzy* c-médias e mapa auto-organizado de Kohonen. Os resultados mostraram que os métodos baseados na Dialética são robustos ao ruído e podem atingir resultados de quantização tão bons quanto aqueles obtidos com o mapa de Kohonen, considerado um quantizador ótimo.

**Palavras-chave** – Segmentação de imagens, k-médias, dialética, otimização, princípio da máxima entropia, computação evolucionária.

**Abstract** – The unsupervised classification has a very important role in the analysis of multispectral images, given its ability to assist the extraction of a priori knowledge of images. Algorithms like k-means and fuzzy c-means has long been used in this task. Computational Intelligence has proven to be an important field to assist in building classifiers optimized according to the quality of the grouping of classes and the evaluation of the quality of vector quantization. Several studies have shown that Philosophy, especially the Dialectical Method, has served as an important inspiration for the construction of new computational methods. This paper presents an evaluation of four methods based on the Dialectics: the Objective Dialectical Classifier and the Dialectical Optimization Method adapted to build a version of k-means with optimal quality indices; each of them is presented in two versions: a canonical version and another version obtained by applying the Principle of Maximum Entropy. These methods were compared to k-means, fuzzy c-means and Kohonen's self-organizing maps. The results showed that the methods based on Dialectics are robust to noise, and quantization can achieve results as good as those obtained with the Kohonen map, considered an optimal quantizer.

**Keywords** – Image segmentation, k-means, dialectics, optimization, principle of maximum entropy, evolutionary computation.

## 1 Introdução

Diversos trabalhos mostram que é possível construir um algoritmo para classificação não supervisionada de imagens baseado no método dialético [1–5]. A principal vantagem dos assim chamados classificadores dialéticos objetivos está na possibilidade de se realizar uma classificação sem saber previamente o número real de classes presente na imagem [6, 7], característica compartilhada com diversas redes neurais construtivas [8–14] e pelo método de agrupamento x-médias [15]. No entanto, também foi visto que é possível usar o método dialético como inspiração para construir métodos de busca e otimização, podendo-se alcançar resultados interessantes quanto à precisão da otimização e ao número de iterações e de avaliações da função dispendidos até se alcançar um determinado valor limiar de otimização [16–18].

Contudo, aproveitando o mesmo princípio da otimização de funções, é possível construir algoritmos de classificação não supervisionada e agrupamento usando métodos evolucionários de otimização, onde os vetores candidatos à solução são construídos com parâmetros do classificador (no caso do mapa de k-médias, esses parâmetros seriam compostos pelas coordenadas dos pesos dos centróides); assim, avaliar uma determinada função objetivo em um determinado ponto equivaleria a realizar uma classificação usando os parâmetros do classificador passados pelo vetor candidato à solução e, após isso, avaliar a qualidade da classificação segundo algum item de avaliação da validade do agrupamento, no caso de se otimizar um método de agrupamento, como o k-médias, por exemplo. Assim, é possível usar técnicas de otimização em substituição ao algoritmo de treinamento do método de agrupamento a otimizar, resultando classificadores com determinadas características otimizadas.

Neste artigo se busca avaliar o uso dos classificadores dialéticos objetivos quanto à qualidade da quantização, usando para tanto índices de fidelidade como medidas indiretas da distorção de quantização, além de índices de validade de agrupamento. Os resultados são comparados com resultados semelhantes obtidos com classificadores baseados no mapa de k-médias, no mapa *fuzzy* c-médias e no mapa auto-organizado de Kohonen.

Também é proposto um novo método de classificação não supervisionada e agrupamento baseado na otimização do mapa de k-médias usando o método dialético de otimização em função de índices de avaliação da validade do agrupamento. Todos os resultados são gerados a partir da classificação de imagens de ressonância magnética sintéticas e comparados entre si, considerando as diversas versões canônica e com entropia maximizada, e com resultados gerados com classificadores baseados no mapa de k-médias padrão, no mapa auto-organizado de Kohonen, e no mapa *fuzzy* c-médias, usando testes estatísticos.

Este artigo está organizado da maneira que segue: na seção 2 são descritos o método dialético de otimização, índices e medidas de validação de agrupamento, o método da máxima entropia, e métodos de otimização do algoritmo de k-médias, além de outras abordagens de segmentação, como o algoritmo *fuzzy* c-médias; nessa mesma seção são apresentadas as imagens utilizadas; na seção 3 são apresentados os resultados obtidos com a abordagem proposta, enquanto na seção 4 são tecidas as conclusões finais.

## 2 Materiais e Métodos

O objetivo principal deste trabalho é mostrar que tanto o classificador dialético objetivo (ODC) quanto o método dialético de busca e otimização (ODM) podem ser utilizados para segmentar imagens multiespectrais, sendo o primeiro um método de classificação e segmentação em si, enquanto o último é utilizado para minimizar uma função de custo de um algoritmo de agrupamento padrão aplicado à segmentação de imagens.

A ideia do uso do ODM na classificação é substituir o algoritmo de treinamento do classificador não supervisionado por um processo de otimização usando o ODM para otimizar o resultado de algum índice de avaliação da qualidade do processo de agrupamento (ou *clustering*) e, em seguida, avaliar o resultado utilizando índices de fidelidade que medem a semelhança entre a imagem quantizada obtida da segmentação e a imagem original.

O método de agrupamento otimizado neste trabalho foi o mapa de k-médias, enquanto as imagens adotadas neste estudo de caso consistem em imagens multiespectrais sintéticas de ressonância magnética. Os resultados em seguida são comparados com aqueles obtidos por outros algoritmos de agrupamento, como o k-médias padrão, o *fuzzy* c-médias e o mapa auto-organizado de Kohonen, quanto à distorção de quantização. Para comparação são utilizados testes estatísticos, como o teste de  $\chi^2$ , para avaliar quão próximos são os resultados de quantização do ponto de vista global, ou seja, considerando diversos resultados para índices de fidelidade diferentes.

### 2.1 Algoritmo de Busca e Otimização

O método dialético objetivo pode ser adaptado para as tarefas de busca e otimização da seguinte maneira [16, 19, 20]:

1. Definem-se a quantidade inicial  $m(0)$  de polos integrantes do sistema dialético  $\Omega(0)$ , juntamente com o número de fases históricas,  $n_P$ , e a duração de cada fase histórica,  $n_H$ . Esses são parâmetros do método dialético de otimização. O número inicial de polos corresponde ao conceito de população inicial das abordagens evolucionárias tradicionais. O número inicial de polos deve ser par, de forma que metade do número de polos seja gerada de forma aleatória, dentro do domínio da função objetivo, enquanto a outra metade é obtida pela geração de polos antítese absoluta. Assim, do ponto de vista da concepção dialética, têm-se um conjunto inicial de polos composto de pares tese-antítese em contradição antagônica (um dos polos é uma tese, enquanto o outro é sua antítese antagônica), gerando maior dinâmica inicial, ou seja, a luta de polos é mais intensa [21, 22]. Já do ponto de vista da computação evolucionária, essa estratégia acelera tipicamente em 10% a convergência do algoritmo de otimização, uma vez que os algoritmos de busca e otimização baseados em computação evolucionária são altamente dependentes da população inicial e, portanto, a probabilidade de encontrar os pontos ótimos utilizando menos iterações pode ser aumentada ao se avaliar tanto um candidato à solução quanto o seu oposto [23]. Contudo, é importante observar que a convergência prematura pode prejudicar a eficácia do algoritmo [23]. Além do mais, a rápida convergência de um algoritmo de otimização pode não implicar em encontrar o ponto ótimo mais rápido para todos os casos [23]. Levando em conta essas considerações, os polos iniciais são definidos da forma que segue:

$$w_{i,j}(0) = \begin{cases} U(r_j, s_j), & 1 \leq i \leq \frac{1}{2}m(0), \\ \check{w}_{i',j}(0), & 1 + \frac{1}{2}m(0) \leq i \leq m(0), \end{cases} \quad (1)$$

$$\check{w}_{i,j} = s_j - w_{i,j} + r_j, \quad (2)$$

para  $i' = i - \frac{1}{2}m(0)$ ,  $1 \leq i \leq m(0)$  e  $1 \leq j \leq n$ , onde  $n$  é a dimensionalidade do problema de otimização,  $U(r_j, s_j)$  é um número aleatório uniformemente distribuído no intervalo  $[r_j, s_j]$  e  $S = [r_1, s_1] \times [r_2, s_2] \times \dots \times [r_n, s_n]$ , desde que  $s_j > r_j$  e  $s_j, r_j \in \mathbb{R}$ .

2. Enquanto não se atinge um máximo de  $n_P$  fases históricas e a força hegemônica histórica não é maior do que um dado limiar superior de força (estimativa inicial do valor máximo da função objetivo),  $f_H(t) < f_{\text{sup}}$  (critério para se considerar o máximo da função objetivo atingido), repete-se:

**Evolução:** Enquanto não se atinge um máximo de  $n_H$  iterações e  $f_H(t) < f_{\text{sup}}$ , os polos são ajustados segundo a seguinte expressão:

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \Delta \mathbf{w}_{C,i}(t) + \Delta \mathbf{w}_{H,i}(t), \quad (3)$$

para

$$\Delta \mathbf{w}_{C,i}(t) = \eta_C(t)(1 - \mu_{C,i}(t))^2(\mathbf{w}_C(t) - \mathbf{w}_i(t)), \quad (4)$$

$$\Delta \mathbf{w}_{H,i}(t) = \eta_H(t)(1 - \mu_{H,i}(t))^2(\mathbf{w}_H(t) - \mathbf{w}_i(t)), \quad (5)$$

$$0 < \eta_C(t) < 1, \quad (6)$$

$$0 < \eta_H(t) < 1, \quad (7)$$

onde  $\eta_C(0) = \eta_H(0) = \eta_0$  e  $0 < \eta_0 < 1$ . Os termos  $\Delta \mathbf{w}_{C,i}(t)$  e  $\Delta \mathbf{w}_{H,i}(t)$  modelam as influências das hegemonias presente e histórica, nesta ordem, sobre o  $i$ -ésimo polo, enquanto  $\eta_C(t)$  e  $\eta_H(t)$  são os respectivos passos de atualização dos polos, atualizados a cada iteração e a cada fase histórica, respectivamente, de forma que

$$\eta_C(t+1) = \alpha \eta_C(t), \quad (8)$$

ao final de cada iteração e

$$\eta_H(t+1) = \alpha \eta_H(t), \quad (9)$$

ao final de cada fase histórica, para  $\alpha < 1$  (tipicamente,  $\alpha = 0,9999$ ). O decrescimento dos passos ao longo do tempo, embora lento, é efetuado para incrementar a busca e, portanto, facilitar a convergência do algoritmo, tal como em algumas redes neurais artificiais, garantindo assim a convergência assintótica do algoritmo [8]. Os termos  $\mu_{C,i}$  e  $\mu_{H,i}$  são chamados de *pertinência presente* e *pertinência histórica*, respectivamente, sendo definidos da forma que segue, baseada nas funções de pertinência da versão clássica do classificador fuzzy c-médias [24–28]:

$$\mu_{C,i}(t) = \left( \sum_{j=1}^{m(t)} \frac{|f(\mathbf{w}_i(t)) - f_C(t)|}{|f(\mathbf{w}_j(t)) - f_C(t)|} \right)^{-1}, \quad (10)$$

$$\mu_{H,i}(t) = \left( \sum_{j=1}^{m(t)} \frac{|f(\mathbf{w}_i(t)) - f_H(t)|}{|f(\mathbf{w}_j(t)) - f_H(t)|} \right)^{-1}, \quad (11)$$

onde  $1 \leq i \leq m(t)$ . Assim, quando  $f(\mathbf{w}_i(t))$  se aproxima de  $f_C(t)$ , o termo  $\mu_{C,i}(t)$  se aproxima de 1, o que aproxima  $\Delta \mathbf{w}_{C,i}(t)$  de 0 e, portanto, torna a influência da correlação de forças presente praticamente nula, evitando alterações no peso devidas à hegemonia presente. Semelhantemente, quando  $f(\mathbf{w}_i(t))$  se aproxima de  $f_H(t)$ , o termo  $\mu_{H,i}(t)$  se aproxima de 1, o que aproxima  $\Delta \mathbf{w}_{H,i}(t)$  de 0, torna a influência da hegemonia histórica praticamente nula.

**Crise Revolucionária:** Na etapa de crise revolucionária são executados os seguintes passos:

- (a) Todas as contradições  $\delta_{i,j}$  são avaliadas; as contradições menores do que uma *contradição mínima*  $\delta_{\min}$  implicam a fusão entre os polos, de forma que

$$\delta_{i,j}(t) > \delta_{\min} \Rightarrow \mathbf{w}_i(t), \mathbf{w}_j(t) \in \Omega(t+1), \quad (12)$$

$$\delta_{i,j}(t) \leq \delta_{\min} \Rightarrow \mathbf{w}_i(t) \in \Omega(t+1). \quad (13)$$

$i \neq j, \forall i, j$  onde  $1 \leq i, j \leq m(t)$  e  $\Omega(t+1)$  é o novo conjunto de polos.

- (b) A partir das contradições avaliadas na etapa anterior, encontram-se aquelas maiores do que uma *contradição máxima*  $\delta_{\max}$ ; essas contradições serão consideradas as *contradições principais* do sistema dialético, sendo os pares de polos envolvidos considerados como *pares tese-antítese*, cujos *polos síntese* passam também a pertencer ao novo conjunto de polos, ou seja:

$$\delta_{i,j}(t) > \delta_{\max} \Rightarrow g(\mathbf{w}_i(t), \mathbf{w}_j(t)) \in \Omega(t+1), \quad (14)$$

para  $i \neq j, \forall i, j$  onde  $1 \leq i, j \leq m(t)$ .

- (c) Adiciona-se o *efeito de crise*, dada a *máxima crise*,  $\chi_{\max}$ , a todos os polos do sistema dialético  $\Omega(t+1)$ , gerando o novo conjunto de polos,  $\Omega(t+2)$ , de forma que  $\mathbf{w}_k(t+2) \in \Omega(t+2)$ , desde que

$$w_{k,i}(t+2) = w_{k,i}(t+1) + \chi_{\max} G(0, 1), \quad (15)$$

para  $1 \leq k \leq m(t+1)$  e  $1 \leq i \leq n$ , onde  $G(0, 1)$  é um número aleatório de distribuição gaussiana com esperança 0 e variância 1.

- (d) Caso o critério de parada ainda não tenha sido atingido (número máximo de fases históricas atingido ou outro critério de parada a ser definido), é gerado um novo conjunto de polos, de forma que:

$$\mathbf{w}_i(t+2) \in \Omega(t+2) \Rightarrow \check{\mathbf{w}}_i(t+2) \in \Omega(t+2), \quad (16)$$

para  $1 \leq i \leq m(t+2)$ , onde  $m(t+2) = 2m(t+1)$ . Logo, o conjunto de polos é ampliado por meio da adição dos polos em antítese antagônica aos polos já existentes. Tal procedimento se dá como uma forma de modelar a concepção dialética de que, ao passar a uma nova fase histórica, o sistema carrega em si também os seus opostos, o que corresponde ao germe de sua potencial transformação em algo novo. Além do mais, esse procedimento repete o que foi assumido quando da inicialização do sistema dialético (geração do conjunto de polos inicial,  $\Omega(0)$ ), podendo acelerar a convergência do algoritmo à solução ótima pela ideia de acrescentar elementos opostos à busca [23].

A figura 1 exibe o fluxograma geral do método dialético objetivo (ODM) adaptado a busca e otimização de forma simplificada. Embora seja relativamente mais complexo do que o fluxograma correspondente aos algoritmos de otimização por enxame de partículas (PSO), o ODM possui a propriedade de ter custo computacional decrescente em cada fase histórica, uma vez que, à medida que a solução se aproxima da solução ótima, os polos se aproximam e, portanto, as contradições entre si diminuem, o que faz com que diversos polos sejam fundidos.

## 2.2 Análise pelo Método da Máxima Entropia

Um método de busca e otimização será tão bom quanto boa for a sua capacidade de descobrir pontos no espaço de busca onde o valor da função objetivo é melhor do que o melhor antes obtido. Outra característica importante é que um algoritmo de busca e otimização será também tão bom quanto for garantida a sua convergência, que deve de preferência convergir para um ponto onde a função objetivo é ótima ou quase ótima (um máximo/mínimo global ou um máximo/mínimo local). Tem-se assim dois critérios básicos de avaliação de um algoritmo de busca e otimização:

1. Capacidade de exploração, ou capacidade de descobrir de novos pontos ótimos (pontos de máximo ou de mínimo, dependendo do problema);
2. Estabilidade e convergência.

Diversos trabalhos buscam analisar a estabilidade e a convergência de um algoritmo de busca (algoritmos genéticos, programação evolucionária e otimização por enxame de partículas, por exemplo) usando desde critérios de estabilidade assintótica derivados do Controle de Processos [29–31], dado que um algoritmo pode ser considerado um processo dinâmico, até a modelagem do algoritmo de busca como uma cadeia de Markov [32–39], uma vez que o espaço de busca é finito e contável, dadas as limitações de representação numérica em sistemas computacionais, sendo o ponto de máximo/mínimo considerado um estado da cadeia.

Neste trabalho buscou-se usar uma abordagem um tanto diferente: foi utilizado o Princípio da Máxima Entropia para gerar uma versão modificada do algoritmo, que será comparada com a versão canônica do método dialético de busca e otimização aqui apresentada. A análise foi focada nos estágios de evolução, o que *a priori* considera fases históricas relativamente grandes. A escolha de focar as etapas de evolução e não de crise revolucionária se deu baseada no fato de que a etapa de crise revolucionária não é qualitativamente diferente do que ocorre nos métodos de busca e otimização baseados em algoritmos genéticos e em programação evolucionária, a não ser pelo processo de fusão de polos semelhantes (baixa contradição entre si) e de geração de novos polos a partir das contradições principais.

Para aplicar o Princípio da Máxima Entropia no método dialético de busca e otimização é necessário partir da expressão de atualização dos pesos dos polos. Para simplificar, aqui se utilizará uma versão onde o passo histórico e o passo contemporâneo são constantes.

$$\mathbf{w}_i(t+1) = \mathbf{w}_i(t) + \Delta \mathbf{w}_{C,i}(t) + \Delta \mathbf{w}_{H,i}(t), \quad (17)$$

para

$$\Delta \mathbf{w}_{C,i}(t) = \eta_C(t)(1 - \mu_{C,i}(t))^2(\mathbf{w}_C(t) - \mathbf{w}_i(t)), \quad (18)$$

$$\Delta \mathbf{w}_{H,i}(t) = \eta_H(t)(1 - \mu_{H,i}(t))^2(\mathbf{w}_H(t) - \mathbf{w}_i(t)), \quad (19)$$

$$\eta_C(t) = \eta_H(t) = \eta_0, \quad (20)$$

$$0 < \eta_0 < 1. \quad (21)$$

Quanto à convergência do algoritmo dialético de busca, é necessário que o processo minimize as seguintes quantidades:

$$E_H = \sum_{i=1}^m \mu_{H,i} |f(\mathbf{w}_i) - f_H|, \quad (22)$$

$$E_C = \sum_{i=1}^m \mu_{C,i} |f(\mathbf{w}_i) - f_C|. \quad (23)$$

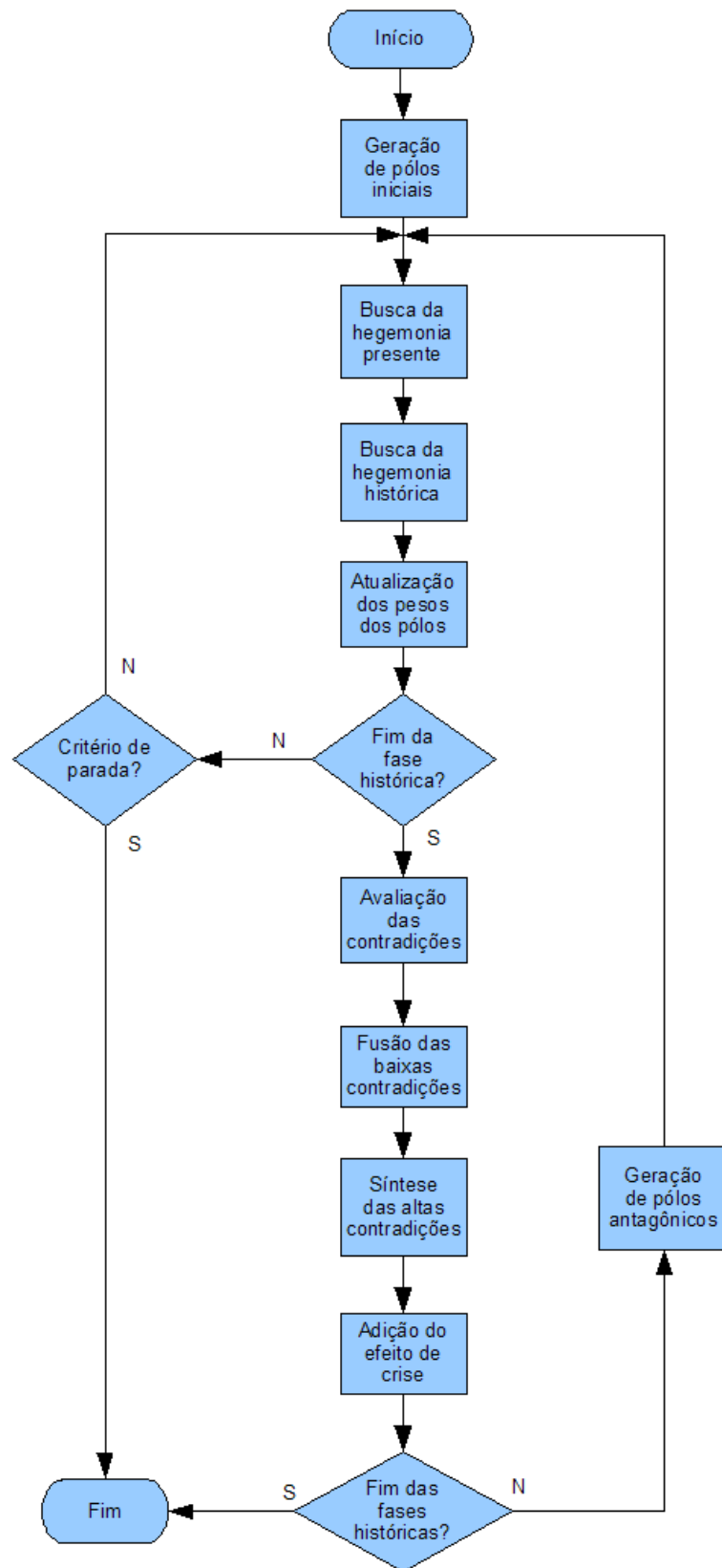


Figura 1: Fluxograma geral do ODM adaptado a busca e otimização

Sabendo-se que as pertinências histórica e contemporânea obedecem à seguintes propriedades:

$$\sum_{i=1}^m \mu_{H,i} = 1, \quad (24)$$

$$0 \leq \mu_{H,i} \leq 1, \quad (25)$$

$$\sum_{i=1}^m \mu_{C,i} = 1, \quad (26)$$

$$0 \leq \mu_{C,i} \leq 1, \quad (27)$$

é possível perceber aqui um caso onde as medidas de incerteza expressas pelas funções de pertinência são na verdade probabilísticas, conforme [40], cabendo portanto o uso da Entropia de Shannon, que neste caso é equivalente à Entropia de Shapley, dado que na verdade  $\mu_{H,i}$  e  $\mu_{C,i}$  expressam as probabilidades  $p_{H,i}$  e  $p_{C,i}$  de o  $i$ -ésimo polo coincidir com o polo hegemônico histórico e com o polo hegemônico contemporâneo, respectivamente. Assim, com base em funções de pertinência, pode-se definir uma *entropia histórica*  $H(\mu_H)$  e uma *entropia contemporânea*  $H(\mu_C)$ , que podem ser maximizadas em paralelo, para maximizar a capacidade exploratória do algoritmo:

$$H(\mu_H) = - \sum_{i=1}^m \mu_{H,i} \ln \mu_{H,i}, \quad (28)$$

$$H(\mu_C) = - \sum_{i=1}^m \mu_{C,i} \ln \mu_{C,i}. \quad (29)$$

Assim, o trabalho se reduz a dois problemas de maximização restritos: a) maximizar  $H(\mu_H)$  para  $E_H$  pequeno; b) maximizar  $H(\mu_C)$  para  $E_C$  pequeno. Isso se assemelha à construção de um algoritmo de agrupamento *fuzzy* c-médias usando o Princípio da Máxima Entropia, utilizando o Método dos Multiplicadores de Lagrange [41]. Assim, são obtidas as seguintes expressões para as funções de pertinência históricas e contemporâneas [40]:

$$\mu_{H,i} = \frac{\exp(-\lambda_H g_{H,i})}{\sum_{j=1}^m \exp(-\lambda_H g_{H,j})}, \quad (30)$$

$$\mu_{C,i} = \frac{\exp(-\lambda_C g_{C,i})}{\sum_{j=1}^m \exp(-\lambda_C g_{C,j})}, \quad (31)$$

ou melhor:

$$\mu_{H,i} = \frac{\exp(-\lambda_H |f(\mathbf{w}_i) - f_H|)}{\sum_{j=1}^m \exp(-\lambda_H |f(\mathbf{w}_j) - f_H|)}, \quad (32)$$

$$\mu_{C,i} = \frac{\exp(-\lambda_C |f(\mathbf{w}_i) - f_C|)}{\sum_{j=1}^m \exp(-\lambda_C |f(\mathbf{w}_j) - f_C|)}, \quad (33)$$

para

$$g_{H,i} = |f(\mathbf{w}_i) - f_H|, \\ g_{C,i} = |f(\mathbf{w}_i) - f_C|,$$

onde  $\lambda_H > 0$  e  $\lambda_C > 0$  são os multiplicadores de Lagrange para as funções de pertinência históricas e contemporâneas, respectivamente. Quando  $\lambda_H \rightarrow 0$ , as funções de pertinência histórica tendem a ser iguais; quando  $\lambda_H \rightarrow +\infty$ , a tendência é de que uma das funções de pertinência histórica tenda a quase 1, enquanto as outras tendem a 0, o que significa que um dos polos tende a convergir rapidamente para o polo hegemônico histórico [42–44]. O mesmo vale para  $\lambda_C$  e as funções de pertinência contemporâneas. Uma vez que os polos hegemônicos não são fixos, dado que se trata de um processo de busca, e nem a quantidade de polos é fixa, devido à etapa de crise revolucionária, as expressões das funções de pertinência se tornam:

$$\mu_{H,i}(t) = \frac{\exp(-\lambda_H(t) |f(\mathbf{w}_i(t)) - f_H(t)|)}{\sum_{j=1}^{m(t)} \exp(-\lambda_H(t) |f(\mathbf{w}_j(t)) - f_H(t)|)}, \quad (34)$$

$$\mu_{C,i}(t) = \frac{\exp(-\lambda_C(t) |f(\mathbf{w}_i(t)) - f_C(t)|)}{\sum_{j=1}^{m(t)} \exp(-\lambda_C(t) |f(\mathbf{w}_j(t)) - f_C(t)|)}. \quad (35)$$

Uma vez que  $f_H(t)$  e  $f_C(t)$  não são fixos, não é possível determinar algebricamente os multiplicadores de Lagrange  $\lambda_H(t)$  e  $\lambda_C(t)$ . Uma alternativa interessante é fazer  $\lambda_H(t)$  e  $\lambda_C(t)$  inicialmente pequenos, para iniciar o algoritmo com uma boa capacidade de exploração, dado que as funções de pertinência passam a ter uma característica mais *fuzzy* (ou seja, com transições mais suaves, por exemplo,  $e^{-\lambda_1 y} > e^{-\lambda_2 y}$ , para  $\lambda_1 < \lambda_2$ ,  $\forall y \in \mathbb{R}$ , e  $\lambda_1, \lambda_2 > 0$ ) e, à medida em que os polos são fundidos,

resultado da convergência para um determinado ponto de mínimo,  $\lambda_H(t)$  e  $\lambda_C(t)$  aumentam. Assim, uma boa escolha poderia ser:

$$\lambda_H(t) = \lambda_C(t) = \frac{1}{m(t)}, \quad (36)$$

o que dá origem às seguintes expressões:

$$\mu_{H,i}(t) = \frac{\exp\left(-\frac{1}{m(t)}|f(\mathbf{w}_i(t)) - f_H(t)|\right)}{\sum_{j=1}^{m(t)} \exp\left(-\frac{1}{m(t)}|f(\mathbf{w}_j(t)) - f_H(t)|\right)}, \quad (37)$$

$$\mu_{C,i}(t) = \frac{\exp\left(-\frac{1}{m(t)}|f(\mathbf{w}_i(t)) - f_C(t)|\right)}{\sum_{j=1}^{m(t)} \exp\left(-\frac{1}{m(t)}|f(\mathbf{w}_j(t)) - f_C(t)|\right)}. \quad (38)$$

Comparando as expressões 34 e 35 com as expressões canônicas 11 e 10, nesta ordem, é possível notar o seguinte: quanto mais próximo  $f(\mathbf{w}_i)$  for de  $f_H$ , maior a pertinência, ou seja:  $\mu_{H,i} \propto 1/d(f(\mathbf{w}_i), f_H)$ , onde  $d: \mathbb{R}^2 \rightarrow \mathbb{R}_+$  é uma medida de distância, que pode também vir a ser simplesmente o módulo da diferença, caso os argumentos sejam escalares. Assim, um caminho intuitivo é simplesmente normalizar  $\mu_{H,i} = K/d(f(\mathbf{w}_i), f_H)$ , onde  $K \in \mathbb{R}_+$ , por  $\sum_{j=1}^m \mu_{H,j} = \sum_{j=1}^m K/d(f(\mathbf{w}_j), f_H)$ , resultando na expressão canônica 11, para  $d(x, y) = |x - y|$ , onde  $x, y \in \mathbb{R}$ . Logo:

$$\lim_{f(\mathbf{w}_i) \rightarrow f_H} \mu_{H,i} = \lim_{f(\mathbf{w}_i) \rightarrow f_H} \frac{1/d(f(\mathbf{w}_i), f_H)}{\sum_{j=1}^m 1/d(f(\mathbf{w}_j), f_H)},$$

que pela Regra de L'Hôpital resulta:

$$\lim_{f(\mathbf{w}_i) \rightarrow f_H} \mu_{H,i} = \lim_{f(\mathbf{w}_i) \rightarrow f_H} \frac{1/d(f(\mathbf{w}_i), f_H)}{1/d(f(\mathbf{w}_i), f_H)} = 1.$$

Uma outra maneira de formular o problema de obter as expressões das funções de pertinência seria fazer

$$\mu_{H,i} \propto \exp[-d(f(\mathbf{w}_i), f_H)],$$

o que é outra maneira de expressar que o aumento da distância entre o valor da função objetivo no  $i$ -ésimo polo e o valor da função no polo hegemônico histórico corresponde à diminuição da pertinência histórica do  $i$ -ésimo polo, evitando que o valor exploda para infinito. Assim, normaliza-se

$$\mu'_{H,i} = K \exp[-d(f(\mathbf{w}_i), f_H)]$$

por

$$\sum_{j=1}^m \mu'_{H,j} = \sum_{j=1}^m K \exp[-d(f(\mathbf{w}_j), f_H)],$$

resultando

$$\mu_{H,i} = \frac{\mu'_{H,i}}{\sum_{j=1}^m \mu'_{H,j}},$$

onde  $K \in \mathbb{R}_+$ , resultando na expressão 32, para  $\lambda_H = 1$ . Análises semelhantes podem ser feitas para as expressões das funções de pertinência contemporâneas  $\mu_{C,i}$ , para  $1 \leq i \leq m$ , sendo obtidos resultados idênticos.

Logo, existem semelhanças intuitivas entre as expressões canônicas das funções de pertinência históricas e contemporâneas e aquelas expressões obtidas pelo uso do Princípio da Máxima Entropia, o que levanta a necessidade de gerar resultados experimentais que possam comparar as duas abordagens usando uma boa quantidade de funções de teste, conforme Wolpert e Macready [45].

As diversas versões do método dialético foram comparadas com outros métodos de classificação não supervisionada e métodos evolutivos de otimização, obtendo resultados bastante razoáveis tanto quanto à tarefa de otimização, quanto ao desempenho da classificação não supervisionada, tendo sido avaliado também o custo computacional da solução [1–7, 16–20, 40].

### 2.3 Aplicações em Classificação e Reconhecimento de Padrões

Os métodos de busca e otimização baseados em computação evolucionária podem ser utilizados em diversas aplicações onde é necessário minimizar uma determinada função custo. Uma dessas aplicações é o uso de computação evolucionária para geração de métodos de agrupamento que, por sua vez, podem ser utilizados para implementar classificadores não supervisionados em aplicações como reconhecimento de padrões e classificação de imagens multiespectrais [46].

Assim, o problema de agrupar os elementos do conjunto  $Z = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{n_Z}\}$ , com  $n_Z$  elementos, em  $n_G$  grupos com centróides  $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_G}\}$  se reduz a minimizar a função

$$J_e = \sum_{i=1}^{n_G} \sum_{\mathbf{z} \in G_i} \frac{d(\mathbf{z}, \mathbf{v}_i)}{n_G n_{G,i}}, \quad (39)$$

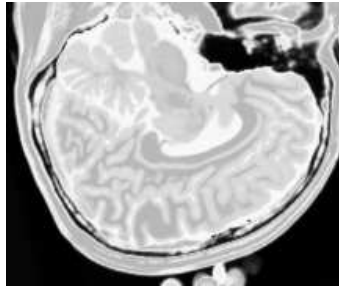
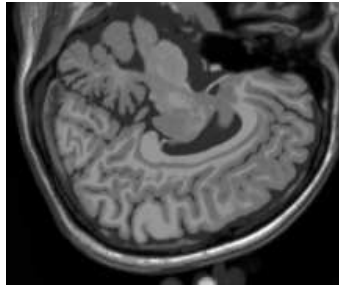


Figura 2: Imagem de RM da fatia 97 ponderada em PD

Figura 3: Imagem de RM da fatia 97 ponderada em  $T_1$ 

onde  $d(\mathbf{z}, \mathbf{v}_i)$  é uma medida de distância entre um vetor  $\mathbf{z}$  da amostra e o centróide do  $i$ -ésimo grupo  $\mathbf{v}_i$ , podendo vir a ser a distância euclidiana, por exemplo, enquanto  $n_{G,i}$  é o número de elementos de  $Z$  agrupados no  $i$ -ésimo grupo,  $G_i$ , e  $J_e$  é uma medida do erro de quantização [46, 47]. Os vetores candidatos a soluções são definidos da forma que segue:

$$\mathbf{x} = (\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_{n_G}^T)^T. \quad (40)$$

Logo, modelando o problema para a solução utilizando o método dialético objetivo, os polos podem assumir a seguinte forma, advinda da modelagem para agrupamento e classificação não supervisionada por otimização por enxame de partículas (PSO) [46, 47]:

$$\mathbf{w}_j = (\mathbf{v}_{j,1}^T, \mathbf{v}_{j,2}^T, \dots, \mathbf{v}_{j,n_G}^T)^T, \quad (41)$$

para  $1 \leq j \leq m$ . Quanto à função a ser minimizada, evidentemente outras funções objetivo podem ser adotadas, uma vez que há diversas formas de se estimar o erro de quantização, como utilizar índices de fidelidade, por exemplo [46, 48].

A escolha da função objetivo a ser minimizada depende de fatores como a velocidade do processo de agrupamento, que no caso do PSO canônico costuma ser muito mais baixa do que a de algoritmos de agrupamento como o k-médias [8], por exemplo, muito embora a qualidade do agrupamento, e portanto do resultado da classificação não supervisionada, utilizando o PSO, é significativamente maior [46, 47].

Também é possível montar métodos de classificação supervisionada tendo por base o conjunto de treinamento e a otimização de funções obtidas a partir da matriz de confusão, tais como a taxa de acerto global e o índice  $\kappa$  de correlação estatística [49, 50].

## 2.4 Imagens Multiespectrais Sintéticas

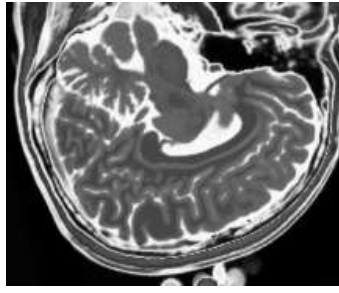
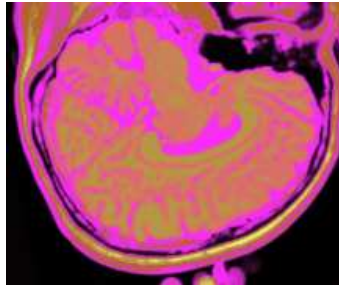
No estudo de caso deste trabalho foram utilizadas 1086 imagens multiespectrais sintéticas de ressonância magnética com três bandas, compostas por 3258 imagens sagitais de 1 mm de espessura, resolução de  $1 \text{ mm}^3$ , ponderadas em densidade de prótons,  $T_1$  e  $T_2$ , para um cérebro humano normal, obtidas por um sistema tomográfico de ressonância magnética com níveis de ruído de 0%, 1%, 3%, 5%, 7% e 9%, e sem inomogeneidades de campo. Essas imagens foram geradas pelo simulador de RM BrainWeb, que permite a geração de imagens sintéticas de RM ponderadas em densidade de prótons,  $T_1$  e  $T_2$  variando o nível de ruído e o percentual de inomogeneidade de campo [51, 52].

As figuras 2 (banda 0), 3 (banda 1) e 4 (banda 2) mostram a fatia 97 de um volume de imagens sagitais com 181 fatias e 0% de ruído, ponderadas em PD (densidade de próton),  $T_1$  e  $T_2$ , enquanto a figura 5 ilustra a composição colorida R0-G1-B2 da mesma fatia. Pode-se notar no topo do crânio, na parte inferior das imagens, a presença de artefatos, que podem ser resultantes de erros no simulador, mas que não são prejudiciais à análise, uma vez que, neste trabalho, não é dada ênfase à análise anatômica.

## 2.5 Quantização Vetorial

A motivação da Teoria da Quantização Vetorial consiste na redução da dimensionalidade da representação dos dados de entrada, visando a compressão de dados. Assim, dado um vetor de entrada  $\mathbf{x}$ , a função  $\mathbf{c}(\mathbf{x})$  é dita um *codificador* de  $\mathbf{x}$ , enquanto  $\mathbf{x}'(\mathbf{c})$  é um *decodificador* de  $\mathbf{c}$ . Na quantização vetorial,  $\mathbf{c}(\mathbf{x})$  transforma  $\mathbf{x}$  de forma a poder ser representado com menos dimensões, ou por meio de uma palavra menor, enquanto  $\mathbf{x}'(\mathbf{c})$  retorna o vetor  $\mathbf{c}$  à mesma base de representação de  $\mathbf{x}$  [8].



Figura 4: Imagem de RM da fatia 97 ponderada em  $T_2$ Figura 5: Composição colorida R0-G1-B2 das imagens da fatia 97 ponderadas em PD,  $T_1$  e  $T_2$ 

Ou seja, dada a transformação não linear  $\Phi : \mathbb{X} \rightarrow \mathbb{A}$ , onde  $\mathbf{x} \in \mathbb{X}$  e  $\mathbb{A}$  é necessariamente um espaço de características discreto, as transformações  $\mathbf{c} = \Phi(\mathbf{x})$  e  $\mathbf{x}' = \Phi^{-1}(\mathbf{c})$  correspondem às tarefas de codificação e decodificação, nesta ordem, em um esquema qualquer de quantização vetorial [8].

Considerando que a entrada  $\mathbf{x}$  possui uma função densidade de probabilidade  $p_{\mathbf{x}}(\mathbf{x})$ , a *distorção esperada* do sistema de quantização será dada pela expressão:

$$D = \frac{1}{2} \int_{\mathbf{x} \in \mathbb{X}} p_{\mathbf{x}}(\mathbf{x}) d(\mathbf{x}, \mathbf{x}') d\mathbf{x}, \quad (42)$$

onde  $d(\mathbf{x}, \mathbf{x}')$  é uma medida de distância entre o vetor de entrada  $\mathbf{x}$  e o vetor reconstruído  $\mathbf{x}'$ , que pode ser tanto a distância euclidiana quanto outra medida de distância, como Mahalanobis, por exemplo. Uma escolha bastante comum é o quadrado da distância euclidiana:  $d(\mathbf{x}, \mathbf{x}') = \|\mathbf{x} - \mathbf{x}'\|^2 = (\mathbf{x} - \mathbf{x}')^T (\mathbf{x} - \mathbf{x}')$  [8]. Assim:

$$D = \frac{1}{2} \int_{\mathbf{x} \in \mathbb{X}} p_{\mathbf{x}}(\mathbf{x}) \|\mathbf{x} - \mathbf{x}'\|^2 d\mathbf{x}. \quad (43)$$

Um sistema de quantização é tão bom quanto menor for a *distorção esperada* e menor for a dimensionalidade de  $\mathbb{A}$ . Obviamente este é um compromisso difícil de atender, pois quanto menor a *distorção*, mais a dimensionalidade de  $\mathbb{A}$  se aproxima da dimensionalidade de  $\mathbb{X}$ , e portanto menor a capacidade de compressão, ao mesmo tempo que, quanto menor a dimensionalidade de  $\mathbb{A}$ , maior a capacidade de compressão sobre o dado reconstruído, mas também maior é a *distorção esperada*.

A quantização *pixel a pixel* de imagens multiespectrais também tem um importante papel quando se deseja comprimir imagens. Na abordagem utilizada neste trabalho, a quantização de imagens multiespectrais coincide com o uso de classificadores não supervisionados baseados em centróides, ou seja, classificadores baseados em métodos de agrupamento de dados (*clustering*). Assim, dado um classificador com  $m$  classes tal que o conjunto de classes seja dado por:

$$\Omega = \{c_1, c_2, \dots, c_m\},$$

onde a cada classe  $c_j$  estão associados um vetor de pesos  $\mathbf{w}_j$  e uma função discriminante  $g_j(\mathbf{x})$ , onde  $1 \leq j \leq m$ , tem-se a seguinte regra, inspirada no critério de decisão de Bayes [49, 50]:

$$g_k(\mathbf{x}) = \max\{g_j(\mathbf{x})\}_{j=1}^m \Rightarrow \mathbf{x}' = \mathbf{w}_k, \quad (44)$$

para  $\mathbf{x} = f(\mathbf{u})$ , onde  $f : S \rightarrow [0, 1]^n$  é uma imagem multiespectral normalizada de  $n$  bandas, e  $f' : S \rightarrow [0, 1]^n$  é a versão reconstruída de  $f$ , dado que  $f'(\mathbf{u}) = \mathbf{x}'$ ,  $\mathbf{u} \in S$  e  $1 \leq k \leq m$ .

Como quantizadores vetoriais foram utilizados neste trabalho: mapas de k-médias, mapas auto-organizados de Kohonen com função de vizinhança gaussiana e com função de vizinhança retangular, mapas *fuzzy* c-médias e classificadores dialéticos objetivos. A dimensionalidade do espaço de características codificadas  $\mathbb{A}$  foi avaliada indiretamente pelo número de classes encontradas nos mapas de classificação, enquanto a *distorção* foi avaliada também indiretamente usando índices de fidelidade global *pixel a pixel*. A classificação não supervisionada também foi avaliada usando medidas de validade de agrupamento.

## 2.6 Índices de Fidelidade

Os principais índices de fidelidade de imagem *pixel a pixel* abordados são o *erro máximo*  $\epsilon_{ME}$  (*Maximum Error*, ME), o *erro médio absoluto*  $\epsilon_{MAE}$  (*Mean Absolute Error*, MAE), o *erro médio quadrático*  $\epsilon_{MSE}$  (*Mean Square Error*, MSE), a *raiz do erro médio quadrático*  $\epsilon_{RMSE}$  (*Root Mean Square Error*, RMSE), o *erro médio quadrático normalizado*  $\epsilon_{NMSE}$  (*Normalized Mean Square Error*, NMSE), a *relação sinal-ruído de pico*  $\epsilon_{PSNR}$  (*Peak Signal to Noise Ratio*, PSNR) e a *relação sinal-ruído*  $\epsilon_{SNR}$  (*Signal to Noise Ratio*, SNR), descritos da forma que segue [48]:

$$\epsilon_{ME} = \max\{\|f(\mathbf{u}) - f'(\mathbf{u})\|\}_{\mathbf{u} \in S}, \quad (45)$$

$$\epsilon_{MAE} = \frac{1}{\#S} \sum_{\mathbf{u} \in S} \|f(\mathbf{u}) - f'(\mathbf{u})\|, \quad (46)$$

$$\epsilon_{MSE} = \frac{1}{\#S} \sum_{\mathbf{u} \in S} \|f(\mathbf{u}) - f'(\mathbf{u})\|^2, \quad (47)$$

$$\epsilon_{RMSE} = \sqrt{\epsilon_{MSE}}, \quad (48)$$

$$\epsilon_{NMSE} = \frac{\sum_{\mathbf{u} \in S} \|f(\mathbf{u}) - f'(\mathbf{u})\|^2}{\sum_{\mathbf{u} \in S} \|f(\mathbf{u})\|^2}, \quad (49)$$

$$\epsilon_{PSNR} = 20 \log_{10} \frac{L_{\max}}{\epsilon_{RMSE}}, \quad (50)$$

$$\epsilon_{SNR} = 10 \log_{10} \frac{1}{\epsilon_{NMSE}}, \quad (51)$$

para  $f : S \rightarrow \{0, 1, \dots, L_{\max}\}^n$ . Para imagens multiespectrais normalizadas  $f : S \rightarrow [0, 1]^n$ , tem-se  $L_{\max} = 1$ .

Existem diversos outros índices para medição de fidelidade de imagens [53–55], como o índice de Wang e Bovik [56]. No entanto, neste trabalho o principal interesse ao usar índices de fidelidade é aproximar a medida da distorção média associada ao processo de quantização. Logo, o principal interesse está em índices baseados em diferenças *pixel a pixel*, o que não é o caso do índice de Wang e Bovik e de outros índices baseados na comparação entre estatísticas globais das imagens [48, 56].

## 2.7 Medidas de Validade de Agrupamento

Os índices de validade de agrupamento (*cluster validity indexes*) correspondem a funções matemáticas e estatísticas para avaliar quantitativamente um algoritmo de agrupamento. Em geral, índices de validade de agrupamento servem para determinar experimentalmente tanto o melhor número de agrupamentos (ou, usando a terminologia de classificação de imagens, o número ideal de classes presentes na imagem) quanto a melhor configuração de agrupamento, ou seja, o melhor particionamento dos dados de entrada. Neste caso, os dados de entrada são os *pixels* da imagem a ser classificada ou segmentada. Assim, uma prática comum é que se façam diversos experimentos até que se encontre o melhor número de classes e a melhor configuração de particionamento [46, 53, 54].

Assim, a classificação ou segmentação de uma imagem é tão boa quanto maior for a *coesão* do processo de agrupamento, ou seja, quanto menor for a maior distância dos pixels ao centróide do grupo ao qual foram associados; o processo de agrupamento também tem sua qualidade avaliada pela *separação* entre os centróides dos grupos: quanto maior for a menor distância entre centróides, melhor o agrupamento. Além disso, quanto menor for o *erro de distorção de quantização*, melhor o processo de agrupamento [46, 53–55].

A seção 2.3 ilustra um índice de avaliação de agrupamento de aplicação geral, que tenta aproximar o erro de distorção, definido da forma que segue, já adaptado para o contexto de segmentação de imagens: o problema de agrupar os *pixels* da imagem  $f : S \rightarrow W^n$ , com altura  $n_H$ , largura  $n_W$  e  $n_B = n$  bandas, resultando  $n_f = n_H \times n_W$  *pixels*, em  $n_G$  grupos (ou classes) com centróides  $V = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{n_G}\}$  se reduz a minimizar a função

$$J_e = \sum_{i=1}^{n_G} \sum_{\forall f(\mathbf{u}) \in G_i} \frac{\|f(\mathbf{u}) - \mathbf{v}_i\|}{n_{G,i}}, \quad (52)$$

onde  $\mathbf{u} \in S$  e  $\|f(\mathbf{u}) - \mathbf{v}_i\|$  é a distância entre o *pixel*  $f(\mathbf{u})$  da imagem  $f : S \rightarrow W^n$  e o centróide do  $i$ -ésimo grupo  $\mathbf{v}_i$ , podendo vir a ser a distância euclidiana, por exemplo, enquanto  $n_{G,i}$  é o número de elementos de  $f$  agrupados no  $i$ -ésimo grupo e  $J_e$  é uma medida do erro de quantização [46, 47]. Os vetores candidatos a soluções são definidos da forma que segue:

$$\mathbf{x} = (\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_{n_G}^T)^T. \quad (53)$$

A coesão do processo de agrupamento é inversamente proporcional à distância máxima intra-agrupamento (*maximum intra-clusters distance*), ou seja, a maior distância entre os *pixels* e os respectivos centróides dos grupos aos quais foram alocados, sendo dada por [46, 53]:

$$d_{\max}(V) = \max_{j=1,2,\dots,n_G} \left\{ \sum_{\forall f(\mathbf{u}) \in G_j} \frac{\|f(\mathbf{u}) - \mathbf{v}_j\|}{n_{G,j}} \right\} = \max_{j=1,2,\dots,n_G} S_j, \quad (54)$$

onde

$$S_j = \sum_{\forall f(\mathbf{u}) \in G_j} \frac{\|f(\mathbf{u}) - \mathbf{v}_j\|}{n_{G,j}}, \quad (55)$$

de onde vem que:

$$J_e = \frac{1}{n_G} \sum_{i=1}^{n_G} S_i. \quad (56)$$

A menor separação dos centróides no processo de agrupamento, ou distância mínima inter-centróides (*minimum inter-clusters distance*) é importante para avaliar a presença de grupos irrelevantes, que em uma análise posterior podem ser eliminados, sendo fundidos a outros mais próximos. Esse índice é dado pela expressão [46, 53]:

$$d_{\min} = \min_{\forall i,j,i \neq j} \{\|\mathbf{v}_i - \mathbf{v}_j\|\}. \quad (57)$$

Pode-se perceber que ao agrupamento ótimo correspondem a maximização da coesão e da separação, junto com a minimização do erro de quantização, o que equivaleria a minimizar  $d_{\max}(V)$ , maximizar  $d_{\min}$  e minimizar  $J_e$  [46]. Omran *et al.* (2005) propõem que é possível construir um novo índice para avaliação do processo de agrupamento de imagens, que pode servir como função objetivo  $f_o : W^{n \times n_G} \rightarrow W$  a ser minimizada por um algoritmo de Computação Evolucionária, a fim de obter um classificador otimizado de acordo com a minimização dessa função objetivo, tendo sido utilizado o índice  $J_o$  que segue, otimizado para gerar um classificador não supervisionado k-médias [46]:

$$f_o(\mathbf{x}_i) = J_o(\mathbf{x}_i) = w_1 d_{\max}(V_i) + w_2 (L_{\max} - d_{\min}(\mathbf{x}_i)) + w_3 J_e(\mathbf{x}_i), \quad (58)$$

onde  $\mathbf{x}_i$  é um candidato à solução e  $V_i$  é o conjunto de centróides associados ao respectivo candidato. Já as grandezas  $w_1$ ,  $w_2$  e  $w_3$  são pesos, que podem ser considerados, em uma primeira abordagem, iguais, ou seja,  $w_1 = w_2 = w_3 = 1/3$ , não privilegiando nenhum aspecto da medida quantitativa da qualidade do agrupamento.

Há diversos outros índices de avaliação do agrupamento, tais como o índice DB, definido por [53]:

$$DB(f, V) = \frac{1}{n_G} \sum_{i=1}^{n_G} \max_{i \neq j} \left\{ \frac{S_i + S_j}{\|\mathbf{v}_i - \mathbf{v}_j\|} \right\}, \quad (59)$$

e o índice de Xie-Beni, dado por [55]:

$$XB(f, V) = \frac{\sum_{i=1}^{n_G} \sum_{\forall \mathbf{u} \in S} \mu_i^2(f(\mathbf{u})) \|f(\mathbf{u}) - \mathbf{v}_i\|^2}{n \min_{i \neq j} \|\mathbf{v}_i - \mathbf{v}_j\|^2}, \quad (60)$$

onde

$$\mu_i(f(\mathbf{u})) = \frac{\|f(\mathbf{u}) - \mathbf{v}_i\|^{2/(1-q)}}{\sum_{j=1}^{n_G} \|f(\mathbf{u}) - \mathbf{v}_j\|^{2/(1-q)}}, \quad (61)$$

para  $q > 1$ .

Omran *et al.* (2005) desenvolveram um algoritmo de agrupamento baseado na otimização do mapa de k-médias segundo a função objetivo definida na expressão 58 utilizando a estratégia de otimização por enxame de partículas (PSO) [46]. Neste trabalho foram gerados resultados utilizando como função objetivo tanto o índice combinado de Omran  $J_o$  quanto apenas o erro de quantização  $J_e$ , otimizando os parâmetros do mapa de k-médias usando o método dialético objetivo de busca e otimização.

## 2.8 Método de Agrupamento de X-Médias

O método de agrupamento de k-médias tem sido um dos métodos mais utilizados para agrupamento geral de dados, dada a sua simplicidade e baixo custo computacional [8, 15]. Contudo, o mapa de k-médias tem como principal inconveniente a necessidade de conhecimento prévio a respeito da quantidade de classes presente no conjunto de dados a ser agrupado [15]. O mapa de x-médias é um método projetado para, dado um conjunto de dados de entrada, encontrar o número de classes que melhor agrupa os dados de entrada, obtendo-se assim um mapa de k-médias para o melhor número de classes, ótimo para um determinado índice de qualidade.

O algoritmo de x-médias é implementado de acordo com o seguinte algoritmo:

1.  $n_G = n_{G,\min}$
2. Enquanto  $n_G \leq n_{G,\max}$ , repita:
  - (a) Alterar parâmetros;
  - (b) Alterar estrutura;

onde os parâmetros de entrada do algoritmo são: o número de classes  $n_G$ , o número mínimo  $n_{G,\min}$ , e o número máximo  $n_{G,\max}$  de classes. O procedimento “alterar parâmetros” consiste em rodar o algoritmo de treinamento k-médias convencional, para um passo inicial  $\eta_0$  e um máximo de  $k_{\max}$  iterações. Já o procedimento “alterar estrutura” consiste em atualizar o número de centroides. Ao final do algoritmo, é retornado o número de classes  $n_G$  que otimiza um determinado índice de qualidade  $J$ .

Neste trabalho foi adotada uma versão modificada do mapa de x-médias, utilizando como índice de qualidade o índice combinado de Omran,  $J_o$ , número de classes variando de 10 a 14, passo inicial  $\eta_0 = 0,1$ , máximo de  $k_{\max} = 200$  iterações. Os resultados foram obtidos usando a imagem multiespectral da fatia 97 do volume, composta pelas imagens ponderadas em densidade de prótons,  $T_1$  e  $T_2$ , para 0% de ruído. Os valores dos índices obtidos são mostrados na tabela 1.

Tabela 1: Índice combinado de Omran em função do número de classes, resultantes do treinamento do x-médias

$n_G$	$J_o$
10	11,47
11	11,13
12	10,38
13	10,45
14	10,45

Pode-se perceber que os resultados para 13 e 14 classes são aproximadamente iguais e correspondem aos menores valores em comparação com os restantes. Assim, pode-se escolher como número de classes presentes na imagem o menor dentre os dois, ou seja, 13 classes. O mapa de k-médias resultante foi utilizado na comparação os outros métodos de classificação não supervisionada, descritos na seção 2.9.

## 2.9 Métodos de Classificação Não Supervisionada

Os métodos de agrupamento, como os mapas auto-organizados de Kohonen, os mapas de k-médias e *fuzzy* c-médias, também podem ser utilizados como métodos de classificação não supervisionada de imagens multiespectrais. Aliás, é fácil mostrar que as tarefas de quantização vetorial e classificação multiespectral são coincidentes, a diferença é que, na primeira, o interesse está nos pesos dos centróides, enquanto na segunda, nos seus índices.

As imagens multiespectrais sintéticas obtidas pelas composições coloridas R0-G1-B2 foram classificadas usando os seguintes métodos, também utilizados para avaliar o desempenho da quantização vetorial:

1. *Mapa auto-organizado de Kohonen (KO)*: 3 entradas, 13 saídas, máximo de 200 iterações, taxa de aprendizado inicial  $\eta_0 = 0,1$ , arquitetura circular, função de distância gaussiana;
2. *Mapa fuzzy c-médias (CM)*: 3 entradas, 13 saídas, máximo de 200 iterações, taxa de aprendizado inicial  $\eta_0 = 0,1$ ;
3. *Mapa de k-médias (KM)*: configuração obtida da aplicação do mapa de x-médias, 3 entradas, 13 saídas, máximo de 200 iterações, taxa de aprendizado inicial  $\eta_0 = 0,1$ ;
4. *Classificador dialético objetivo canônico (ODC-CAN)*: 14 polos iniciais, 2 fases históricas de 150 iterações cada fase, passo histórico inicial  $\eta_0 = 0,1$ , força mínima de 5%, contradição mínima de 0,01, contradição máxima de 0,98, máxima crise de 35%, até 12 polos. Depois de todas as fases históricas, o processo de treinamento finalizou com 13 polos;
5. *Classificador dialético objetivo com entropia máxima (ODC-PME)*: 14 polos iniciais, 2 fases históricas de 150 iterações cada fase, passo histórico inicial  $\eta_0 = 0,1$ , força mínima de 5%, contradição mínima de 0,01, contradição máxima de 0,98, máxima crise de 35%, até 12 polos. Depois de todas as fases históricas, o processo de treinamento finalizou com 13 polos. Aqui foi aplicado o Princípio da Máxima Entropia para obter novas expressões para as funções de anticontração, semelhantemente ao *fuzzy* c-médias com máxima entropia, resultando praticamente as mesmas expressões das funções de pertinência do *fuzzy* c-médias com máxima entropia [41].

O número de entradas é igual ao número de bandas da imagem, uma vez que a classificação é *pixel a pixel*, sendo portanto o *pixel* a entrada da rede. Já o número de saídas foi definido como sendo o número de classes presentes na imagem, que por sua vez é o número de elementos anatômica e fisiologicamente diferentes presentes na imagem (11 classes) mais duas classes para alocar diferentes populações de ruído [51,52]. Os demais parâmetros foram definidos empiricamente. Note que o número final de polos dos classificadores dialéticos coincidiu com o número de classes obtido pelo mapa de x-médias: 13 classes.

## 2.10 Classificadores Otimizados

As imagens multiespectrais sintéticas obtidas pelas composições coloridas R0-G1-B2 foram classificadas usando os seguintes métodos baseados na otimização do mapa de k-médias, considerando 13 classes, 20 polos iniciais, máximo de 10 fases históricas, com duração máxima de 20 gerações cada fase histórica, passos iniciais (histórico e contemporâneo) de 0,99, contradição mínima de 0,1, contradição máxima de 0,9, efeito de crise máximo de 0,9, valor limite de 0,01:

1. *EQ-CAN-KM*: k-médias otimizado pelo método dialético canônico, de acordo com o erro de quantização  $J_e$ ;

2. *IC-CAN-KM*: k-médias otimizado pelo método dialético canônico, de acordo com o índice combinado de Omran  $J_o$ ;
3. *EQ-PME-KM*: k-médias otimizado pelo método dialético com entropia maximizada, de acordo com o erro de quantização  $J_e$ ;
4. *IC-PME-KM*: k-médias otimizado pelo método dialético com entropia maximizada, de acordo com o índice combinado de Omran  $J_o$ .

Os parâmetros iniciais foram definidos de forma semelhante ao que foi descrito na seção anterior. Os parâmetros do método dialético foram definidos empiricamente.

Uma vez que a velocidade de convergência e os resultados de qualquer algoritmo de busca e otimização são altamente dependentes da inicialização do algoritmo, foram utilizadas 5 iterações do algoritmo k-médias padrão para geração do conjunto inicial de polos, de acordo com o critério de geração da população inicial proposto por Saha e Bandyopadhyay (2005), que evita buscas iniciais totalmente aleatórias, uma vez que, embora apenas 5 iterações não sejam suficientes para garantir a convergência do algoritmo k-médias, e nem essa convergência é desejada, uma vez que se está utilizando um algoritmo de otimização em substituição ao algoritmo de treinamento, essas 5 iterações iniciais ajudam o algoritmo de otimização a iniciar com polos que podem estar na vizinhança da solução ótima, sendo este portanto um critério melhor do que inicializar aleatoriamente, evitando buscas totalmente aleatórias e variações quanto à convergência do algoritmo [57].

### 3 Resultados e Discussões

Para facilitar a avaliação dos resultados obtidos, estes foram divididos em duas categorias: resultados gerados usando classificadores dialéticos objetivos; e resultados gerados usando mapa de k-médias otimizado pelo método dialético de otimização, em suas versões canônica e com maximização da entropia, em função do índice combinado de Omran e do erro de quantização. Tanto a qualidade da quantização quanto a da segmentação são avaliadas usando índices de validade de agrupamento e índices de fidelidade da imagem quantizada em relação à imagem original.

#### 3.1 Classificadores Dialéticos Objetivos

A figura 6 mostra os resultados de classificação, enquanto a figura 7 exibe os resultados de quantização para a imagem sem ruído da fatia 97, figura 5, usando os métodos KO, CM, KM, ODC-PME e ODC-CAN. Esses resultados ilustram qualitativamente as diferenças entre os métodos de classificação e quantização, dado que a fatia 97 possui todas as 13 classes presentes na análise [51].

A tabela 2 mostra os resultados da avaliação dos métodos de classificação não supervisionada quanto à quantização vetorial, usando os índices de fidelidade  $\epsilon_{ME}$ ,  $\epsilon_{MAE}$ ,  $\epsilon_{MSE}$ ,  $\epsilon_{RMSE}$  e  $\epsilon_{PSNR}$ , considerando todas as 181 fatias com 3 bandas ( $DP$ ,  $T_1$  and  $T_2$ ), para os métodos KO, CM, KM, ODC-PME e ODC-CAN e 0% de ruído. Já os gráficos das figuras 8, 9, 10 e 11 mostram os resultados em função do nível percentual de ruído para os diversos métodos, para um total de 6 volumes de 181 fatias de 3 bandas, totalizando 1086 imagens coloridas, ou 3258 imagens em níveis de cinza.

As tabelas 3 e 5 mostram o resultado das comparações entre os pares de média e desvio médio das medidas de  $\epsilon_{ME}$ ,  $\epsilon_{MAE}$ ,  $\epsilon_{MSE}$ ,  $\epsilon_{RMSE}$  e  $\epsilon_{PSNR}$ , para os métodos KO, CM, KM, ODC-CAN e ODC-PME, de acordo com teste F de hipótese nula, enquanto as tabelas 4 e 6 mostram comparações entre os métodos ODC-PME e ODC-CAN, e os métodos KO, CM e KM, respectivamente, usando o teste de  $\chi^2$  para medir o grau de dependência (considerado aqui como grau de similaridade) entre os resultados dos métodos ODC-PME e ODC-CAN e os resultados dos outros métodos; as sequências de valores observados foram construídas usando os pares de média e desvio médio dos índices de fidelidade  $\epsilon_{ME}$ ,  $\epsilon_{MAE}$ ,  $\epsilon_{RMSE}$  e  $\epsilon_{PSNR}$ . Todos os testes foram executados considerando um intervalo de confiança de 95%. As figuras 12 e 13 mostram o resultado da similaridade medida entre os métodos ODC-PME e ODC-CAN e os outros métodos, respectivamente, usando o teste de  $\chi^2$ , em função do ruído percentual.

Tabela 2: Resultados de quantização expressos pelos índices de fidelidade, para 0% de ruído

	KO	CM	KM	ODC-PME	ODC-CAN
$\epsilon_{ME}$	50 ± 5	106 ± 15	84 ± 22	50 ± 3	46 ± 6
$\epsilon_{MAE}$	10 ± 3	10 ± 2	13 ± 4	12 ± 2	10 ± 2
$\epsilon_{MSE}$	186 ± 52	258 ± 81	347 ± 261	249 ± 60	169 ± 46
$\epsilon_{RMSE}$	13 ± 2	16 ± 31	18 ± 5	15 ± 2	13 ± 2
$\epsilon_{PSNR}$	26 ± 2	24 ± 2	24 ± 3	25 ± 1	26 ± 2

Tabela 3: Graus de similaridade entre o método ODC-PME e os métodos KO, CM e KM, de acordo com o teste de hipótese nula F, para 0% de ruído

	$\mu(\epsilon_{ME})$	$\mu(\epsilon_{MAE})$	$\mu(\epsilon_{RMSE})$	$\mu(\epsilon_{PSNR})$
ODC-PME-KO	0,99	0,86	0,89	1,00
ODC-PME-CM	0,61	0,86	0,89	1,00
ODC-PME-KM	0,83	0,86	0,89	1,00

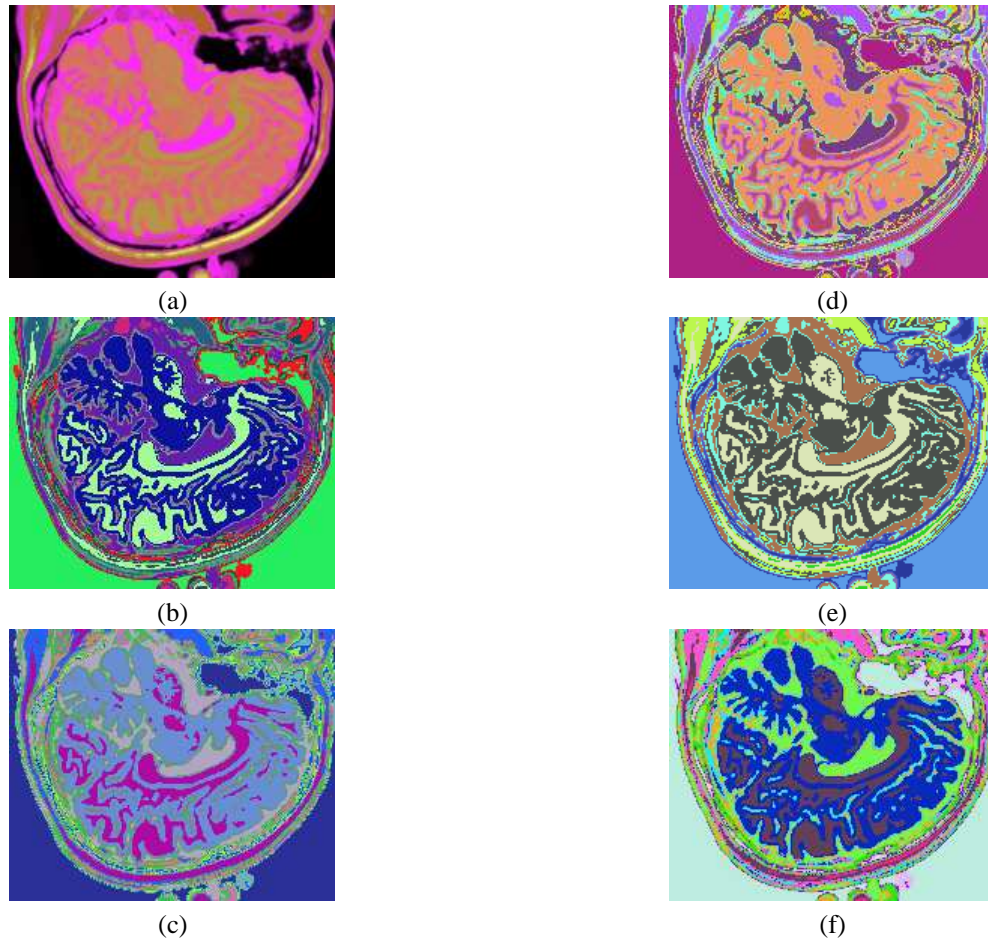


Figura 6: Composição colorida R0-G1-B2 das imagens da fatia 97 ponderadas em PD,  $T_1$  e  $T_2$  (a) e resultados de classificação usando os métodos KO (b), CM (c), KM (d), ODC-PME (e) e ODC-CAN (f)

Tabela 4: Graus de similaridade entre o método ODC-PME e os métodos KO, CM e KM, de acordo com o teste de  $\chi^2$ , em função do ruído

pn (%)	ODC-PME-KO	ODC-PME-CM	ODC-PME-KM
0	0,956111	0,000000	0,000012
1	0,970435	0,000060	0,079578
3	0,999864	0,112552	0,986766
5	0,999971	0,994726	0,998794
7	1,000000	0,995641	0,992783
9	0,994829	0,998837	0,998837

Tabela 5: Graus de similaridade entre o método ODC-CAN e os métodos KO, CM e KM, de acordo com o teste de hipótese nula F, para 0% de ruído

	$\mu(\epsilon_{ME})$	$\mu(\epsilon_{MAE})$	$\mu(\epsilon_{RMSE})$	$\mu(\epsilon_{PSNR})$
ODC-CAN-KO	0,93	1,00	1,00	0,97
ODC-CAN-CM	0,54	1,00	1,00	0,97
ODC-CAN-KM	0,74	1,00	1,00	0,97

Tabela 6: Graus de similaridade entre o método ODC-CAN e os métodos KO, CM e KM, de acordo com o teste de  $\chi^2$ , em função do ruído

pn (%)	ODC-CAN-KO	ODC-CAN-CM	ODC-CAN-KM
0	0,994985	0,000000	0,000005
1	0,999750	0,000069	0,044842
3	0,993934	0,100309	0,962277
5	0,956870	0,915891	0,954338
7	0,992681	0,998602	0,976113
9	0,747098	0,997809	0,997809

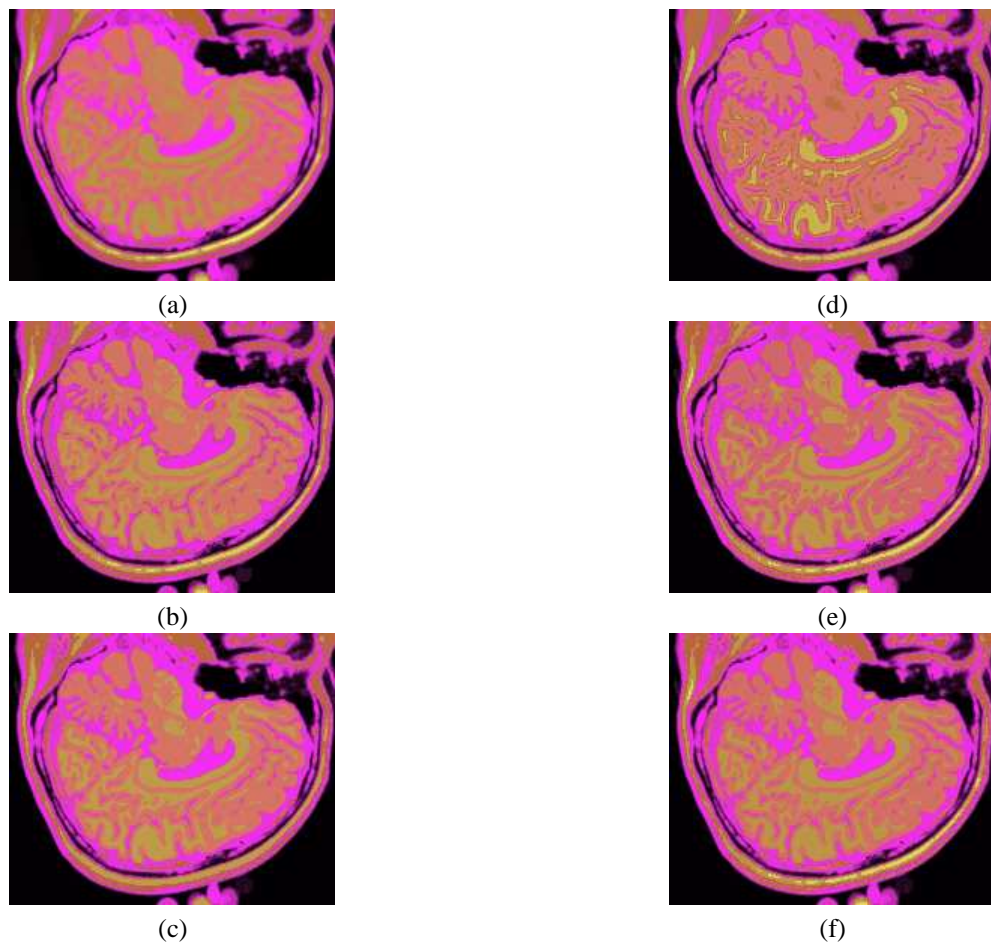


Figura 7: Composição colorida R0-G1-B2 das imagens da fatia 97 ponderadas em PD,  $T_1$  e  $T_2$  (a) e resultados de quantização usando os métodos KO (b), CM (c), KM (d), ODC-PME (e) e ODC-CAN (f)



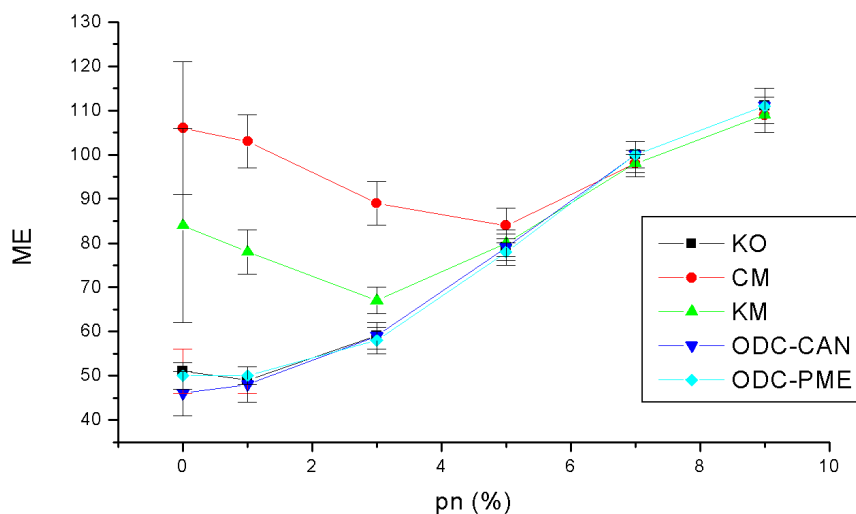


Figura 8: Resultados de  $\epsilon_{ME}$  em função do ruído percentual (ODC)

Observando o gráfico da figura 8 pode-se notar que os métodos ODC-CAN, ODC-PME e KO são os que atingem os melhores resultados para  $\epsilon_{ME}$ , para 0% de ruído. No entanto, os resultados vão se aproximando à medida em que o nível de ruído aumenta, e acabam por se tornar totalmente coincidentes a partir de 5% de ruído. Já os gráficos das figuras 9, 10 e 11 mostram que os resultados de  $\epsilon_{MAE}$ ,  $\epsilon_{RMSE}$  e  $\epsilon_{PSNR}$  obtidos pelos métodos de classificação utilizados não são distinguíveis, não sendo possível comparar os métodos entre si usando os índices de fidelidade citados. Conseqüentemente, a comparação dos métodos índice a índice não é muito conclusiva quanto à classificação dos métodos de segmentação, como mostram os gráficos das figuras 8, 9, 10 e 11, para todos os níveis de ruído, e os testes F exibidos nas tabelas 3 e 5, que comparam os métodos ODC-PME e ODC-CAN com os métodos KO, CM e KM, para 0% de ruído.

Uma vez que as comparações índice a índice não são muito conclusivas, foi utilizado o teste de  $\chi^2$  para fazer comparações globais entre os métodos ODC-PME e ODC-CAN e os métodos KO, CM e KM. A tabela 4 mostra os resultados do teste de  $\chi^2$  entre o método ODC-PME e os outros métodos, mostrando que, para 0% de ruído, os métodos ODC-PME e KO são idênticos. O gráfico da figura 12 ilustra os resultados para os níveis de ruído de 0% a 9%, mostrando que os métodos ODC-PME e KO são praticamente iguais quanto aos resultados dos índices de fidelidade, apesar de todos os métodos atingirem resultados idênticos a partir de 5% de ruído.

A tabela 6 mostra os resultados do teste de  $\chi^2$  entre o método ODC-CAN e os outros métodos, mostrando que, para 0% de ruído, os métodos ODC-CAN e KO são similares. O gráfico da figura 13 ilustra os resultados para os níveis de ruído de 0% a 9%, mostrando que os métodos ODC-CAN e KO são muito similares quanto aos resultados dos índices de fidelidade até 7% de ruído, apesar de todos os métodos atingirem resultados semelhantes a partir de 5% de ruído, enquanto a partir de 7% de ruído os métodos ODC-CAN e KO passam a se diferenciar, comportamento bem diferente daquele observado no método ODC-PME quando comparado com o método KO.

Esses resultados mostram os classificadores dialéticos objetivos em sua versão canônica e na versão com entropia maximizada podem atingir resultados de segmentação e quantização similares àqueles obtidos usando mapas auto-organizados de Kohonen, principalmente quando se trata da versão com entropia maximizada, o que é uma característica muito importante, uma vez que os mapas auto-organizados de Kohonen são quantizadores vetoriais ótimos [8], com a vantagem de os classificadores dialéticos objetivos terem número de classes adaptável, não sendo necessário conhecer o número de classes presente na imagem, bastando inicializar a classificação com um número de classes maior do que o esperado.

### 3.2 Classificadores K-Médias Otimizados pelo Método Dialético

Os gráficos das figuras 14, 15, 16 e 17 mostram os resultados das medidas dos índices de fidelidade  $\epsilon_{ME}$ ,  $\epsilon_{MAE}$ ,  $\epsilon_{RMSE}$  e  $\epsilon_{PSNR}$ , enquanto os gráficos das figuras 18 e 19 exibem os resultados do erro de quantização,  $J_e$ , e do índice de Omran,  $J_o$ , para os métodos KO, CM, KM, IC-CAN, EQ-CAN, IC-PME e EQ-PME, em função dos níveis percentuais de ruído, para os 6 volumes de 181 fatias de 3 bandas. A figura 20 mostra os resultados de classificação, enquanto a figura 21 exhibe os resultados de quantização para a imagem sem ruído da fatia 97, figura 5, usando os métodos IC-CAN-KM, IC-PME-KM, EQ-CAN-KM e EQ-PME-KM. É possível perceber a partir das imagens que a introdução da otimização dialética permitiu ao k-médias identificar diferentes classes dentro do fundo da imagem.

O gráfico da figura 14 mostra que o método EQ-PME, ou seja, o mapa de k-médias otimizado pelo método dialético de



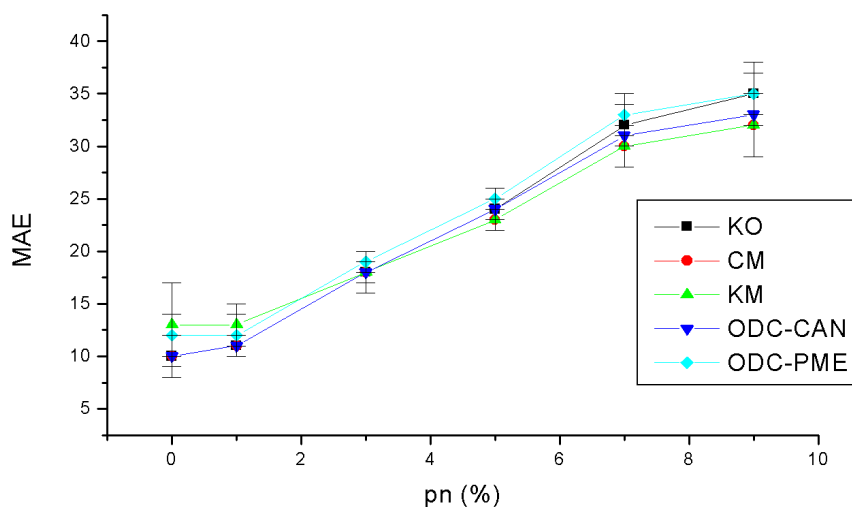


Figura 9: Resultados de  $\epsilon_{MAE}$  em função do ruído percentual (ODC)

entropia maximizada em função do erro de quantização  $J_e$ , foi o que apresentou melhor resultado quanto ao índice de fidelidade  $\epsilon_{ME}$ .

Já o gráfico da figura 15 mostra que os melhores resultados do índice de fidelidade  $\epsilon_{MAE}$  são aqueles obtidos pelos métodos KO e CM, de 0% a 3% de ruído percentual, mas a partir de 3% o método EQ-PME passa a apresentar os melhores resultados, mostrando que o mapa auto-organizado de Kohonen e o mapa *fuzzy* c-médias são pouco robustos em relação ao ruído, enquanto o mapa de k-médias otimizado pelo método dialético com máxima entropia em função do erro de quantização tem desempenho menos sensível aos níveis de ruído; esse resultado, contudo, não é muito diferente do resultado obtido pelo uso do mapa de k-médias.

O gráfico da figura 16 mostra que o mapa auto-organizado de Kohonen apresentou os melhores resultados do índice de fidelidade  $\epsilon_{RMSE}$  para níveis de ruído percentual de 0% a 3%. Contudo, a partir de 3% os resultados obtidos pelo método KO simplesmente não são mais distinguíveis dos resultados para os outros métodos. Já gráfico da figura 17 ilustra que os resultados da aplicação de todos os métodos não são estatisticamente distintos entre si de acordo com o índice de fidelidade  $\epsilon_{RMSE}$ , e isso levando em conta todos os níveis de ruído percentual, mas há uma leve indicação de superioridade dos resultados obtidos com o método de k-médias otimizado usando o método dialético com entropia otimizada em função do erro de quantização, EQ-PME.

Entretanto, o gráfico da figura 18 ilustra que, embora os resultados de 0% a 1% de ruído mostrem que os resultados não são distinguíveis entre si quanto ao erro de quantização, a partir de 1% de ruído os valores do erro de quantização para o método EQ-PME são bem menores do que as medidas de erro de quantização para os outros métodos, mostrando que a otimização do mapa de k-médias usando o método dialético com entropia maximizada em função do erro de quantização tem resultados qualitativamente superiores àqueles obtidos com os outros métodos, conforme também a figura 21.

As diferenças entre os métodos KO, CM e KM e os métodos EQ-CAN, EQ-PME, IC-CAN e IC-PME são mais gritantes quanto a avaliação é feita usando o índice combinado de Omran, como ilustra a figura 18: todos os métodos baseados na otimização do mapa de k-médias por versões do método dialético retornam medidas de  $J_o$  que são pelo menos metade das mesmas medidas obtidas pelos outros métodos.

No entanto, da figura 18 também fica claro que, embora todos os métodos baseados na otimização pelo método dialético retornem medidas de  $J_o$  muito melhores que aquelas obtidas usando KO, CM e KM, o método EQ-PME é o que apresenta o melhor resultado dentre todos, destacando-se dos outros métodos a partir de 1% de ruído percentual, ou seja: o método baseado no mapa de k-médias otimizado usando o método dialético de entropia maximizada em função do erro de quantização  $J_e$ . Esse resultado não deixa de ser interessante e curioso, pois mostra que o índice de Omran  $J_o$ , no caso das imagens utilizadas, é dominado pelo erro de quantização  $J_e$ , dado que a otimização em função do índice de Omran usando os métodos dialéticos canônico (IC-CAN) e com entropia maximizada (IC-PME) são idênticos e qualitativamente inferiores, assim como os resultados obtidos usando o método dialético canônico em função de  $J_e$  (EQ-CAN). Isso também indica que a aplicação do Princípio da Máxima Entropia diferenciou sensivelmente o algoritmo de sua versão canônica, tendo acelerado a convergência para a otimização em função do erro de quantização, como bem atestam os resultados.

Esses resultados, combinados àqueles apresentados anteriormente, mostram que, para a aplicação ilustrada neste trabalho, o método EQ-CAN foi superior aos outros métodos de classificação não supervisionada e quantização apresentados, tanto quando são utilizados índices de fidelidade quanto índices de validade do agrupamento, o que mostra que, para as imagens utilizadas, tanto existe um predomínio do erro de quantização  $J_e$  sobre o índice combinado de Omran  $J_o$  (onde este último, por sua vez,

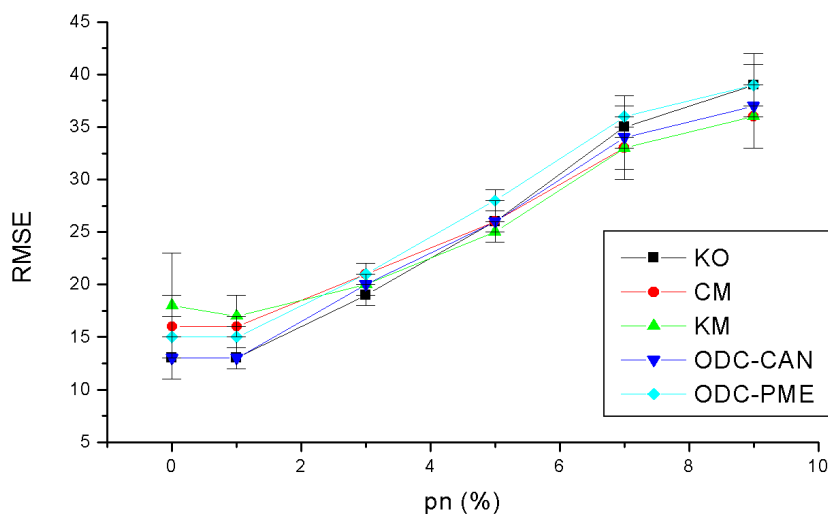


Figura 10: Resultados de  $\epsilon_{RMSE}$  em função do ruído percentual (ODC)

contém o primeiro conforme expressões 58 e 52), quanto existe uma forte influência da aplicação do Princípio da Máxima Entropia ao método dialético de busca e otimização, que nesta aplicação ajudou o método a ter sua convergência acelerada ou até mesmo a vencer um mínimo local.

#### 4 Conclusões

Este trabalho mostrou duas aplicações do método dialético: classificadores dialéticos objetivos e classificadores baseados na otimização dialética. Os classificadores dialéticos objetivos são uma implementação direta do método dialético como definido neste trabalho para problemas de classificação e agrupamento. Foi visto que esses classificadores podem chegar a atingir resultados tão bons quanto os resultados atingidos usando o mapa auto-organizado de Kohonen, considerado um método ótimo para quantização vetorial, com a vantagem de não ser necessário especificar o número exato de classes no início do processo de treinamento, fazendo com que o método seja executado até que se atinja um bom número de classes, necessariamente menor ou igual ao número inicial de classes definidas no início do treinamento.

Já os classificadores baseados na otimização dialética mostraram que é possível obter classificadores otimizados segundo algum critério usando as diversas versões do método dialético e, com eles, conseguir resultados de quantização e de fidelidade superiores àqueles obtidos utilizando o método com parâmetros gerados pela utilização de seu algoritmo próprio de treinamento. Esses classificadores otimizados, quando comparados aos que não fazem uso de otimização e sim utilizam seus próprios algoritmos de treinamento, podem ser uma importante alternativa para evitar com mais facilidade que o processo fique preso em mínimos locais, dada a boa capacidade do método dialético de evitar mínimos locais, como se percebe das situações onde a função objetivo é multimodal, com destaque para a versão com entropia maximizada.

Quanto aos classificadores dialéticos objetivos, os resultados mostram que os classificadores dialéticos e o mapa de Kohonen são os que atingem os melhores resultados para 0% de ruído. No entanto, os resultados obtidos pelos métodos de classificação utilizados não são distinguíveis, não sendo possível comparar os métodos entre si usando índices de fidelidade. Utilizando o teste de  $\chi^2$  entre os classificadores dialéticos e os outros métodos, foi mostrado que, para 0% de ruído, os classificadores dialéticos e o mapa de Kohonen são praticamente idênticos (similaridade acima de 96%), principalmente para a versão com entropia maximizada. Esses resultados mostram que esses classificadores podem atingir resultados de segmentação e quantização similares àqueles obtidos usando mapas auto-organizados de Kohonen, o que é muito importante, já que os mapas de Kohonen são quantizadores vetoriais ótimos, com a vantagem de os classificadores dialéticos objetivos terem número de classes adaptável, não sendo necessário conhecer o número de classes presente na imagem.

Quanto aos mapas de k-médias otimizados pelos métodos dialéticos de otimização, os resultados mostram que o mapa de k-médias otimizado pelo método dialético de entropia maximizada em função do erro de quantização foi o que apresentou melhores resultados quanto aos índices de fidelidade a partir de 3% de ruído. A partir de 3% os resultados obtidos com o mapa de Kohonen não são mais distinguíveis. A partir de 1% de ruído os valores do erro de quantização para o mapa de k-médias otimizado pelo método dialético de entropia maximizada em função do erro de quantização são bem menores do que as medidas de erro de quantização para os outros métodos, mostrando que tem resultados qualitativamente superiores àqueles obtidos com os outros métodos.

As diferenças entre os métodos são realçadas ao se usar o índice combinado de Omran. Todos os métodos baseados na

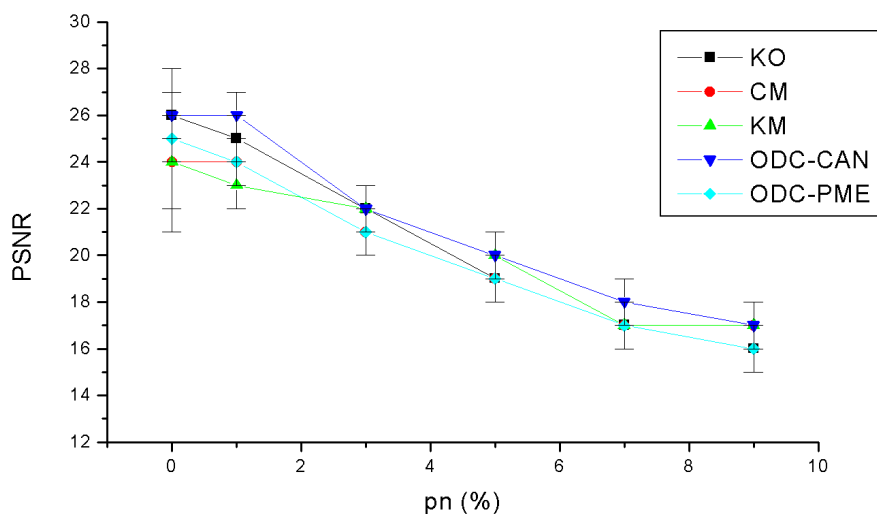


Figura 11: Resultados de  $\epsilon_{PSNR}$  em função do ruído percentual (ODC)

otimização pelo método dialético resultaram em medidas muito melhores que aquelas obtidas usando os outros métodos. Isso indica que a aplicação do Princípio da Máxima Entropia diferenciou sensivelmente o algoritmo de sua versão canônica, tendo acelerado a convergência para a otimização em função do erro de quantização, como bem atestam os resultados. Entretanto, para a aplicação ilustrada neste trabalho, o método EQ-CAN foi superior aos outros métodos apresentados, tanto quando são utilizados índices de fidelidade quanto índices de validade do agrupamento.

Como trabalho futuro propõe-se a avaliação comparativa do classificador dialético objetivo, em suas versões canônica e otimizada pelo Princípio da Máxima Entropia, e do método de agrupamento de  $x$ -médias quanto ao número de classes presente no conjunto de dados de entrada. Como estudos de caso, poderiam ser utilizadas imagens de natureza diversa, com números de classes conhecidos. Outros trabalhos futuros podem envolver o agrupamento de dados numéricos multidimensionais, compostos de classes linearmente e não linearmente separáveis, comparando as diversas versões do classificador dialético com o  $x$ -médias, o  $k$ -médias, o *fuzzy*  $c$ -médias e outros métodos de agrupamento, com ou sem ajuste automático do número de classes.

## Referências

- [1] W. P. Santos, F. M. Assis, R. E. Souza, P. B. Mendes, H. S. S. Monteiro and H. D. Alves. “Dialectical Non-Supervised Image Classification”. In *IEEE Congress on Evolutionary Computation (CEC 2009)*, pp. 2480–2487, Trondheim, 2009. CIS-IEEE.
- [2] W. P. Santos, F. M. Assis, R. E. Souza and P. B. Santos Filho. “Dialectical Classification of MR Images for the Evaluation of Alzheimer’s Disease”. In *Recent Advances in Biomedical Engineering*, edited by G. R. Naik, pp. 241–250, Vukovar, 2009. In-Teh.
- [3] W. P. Santos, F. M. Assis, R. E. Souza, P. B. Mendes, H. S. S. Monteiro and H. D. Alves. “A Dialectical Method to Classify Alzheimer’s Magnetic Resonance Images”. In *Evolutionary Computation*, edited by W. P. Santos, pp. 473–486, Vukovar, 2009. In-Teh.
- [4] W. P. Santos, F. M. Assis, R. E. Souza, A. C. T. C. Albuquerque and M. L. B. Simas. “A Monospectral Approach for fMRI Analysis using Kohonen Self-Organized Networks and Objective Dialectical Classifiers”. *International Journal of Innovative Computing and Applications*, vol. 1, no. 4, pp. 260–273, 2008.
- [5] W. P. Santos, F. M. Assis, R. E. S. P. B. Santos Filho and F. B. Lima Neto. “Dialectical Multispectral Classification of Diffusion-Weighted Magnetic Resonance Images as an Alternative to Apparent Diffusion Coefficients Maps to Perform Anatomical Analysis”. *Computerized Medical Imaging and Graphics*, vol. 33, no. 6, pp. 442–460, 2009.
- [6] W. P. Santos, R. E. Souza, P. B. Santos Filho, F. B. Lima Neto and F. M. Assis. “A Dialectical Approach for Classification of DW-MR Alzheimer’s Images”. In *IEEE World Congress on Computational Intelligence (WCCI 2008)*, pp. 1728–1735, Hong Kong, China, 2008. CIS-IEEE.

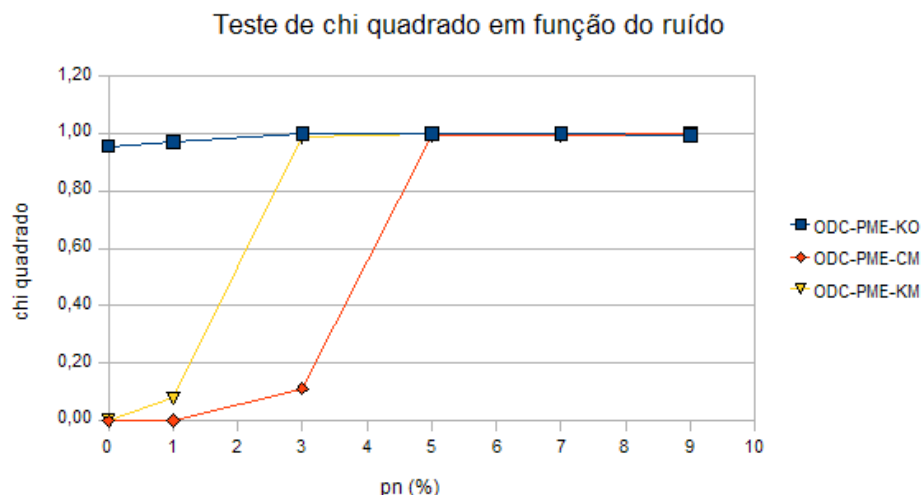


Figura 12: Graus de similaridade do método ODC-PME em relação aos métodos KO, CM e KM, usando o teste de  $\chi^2$ , em função do ruído percentual

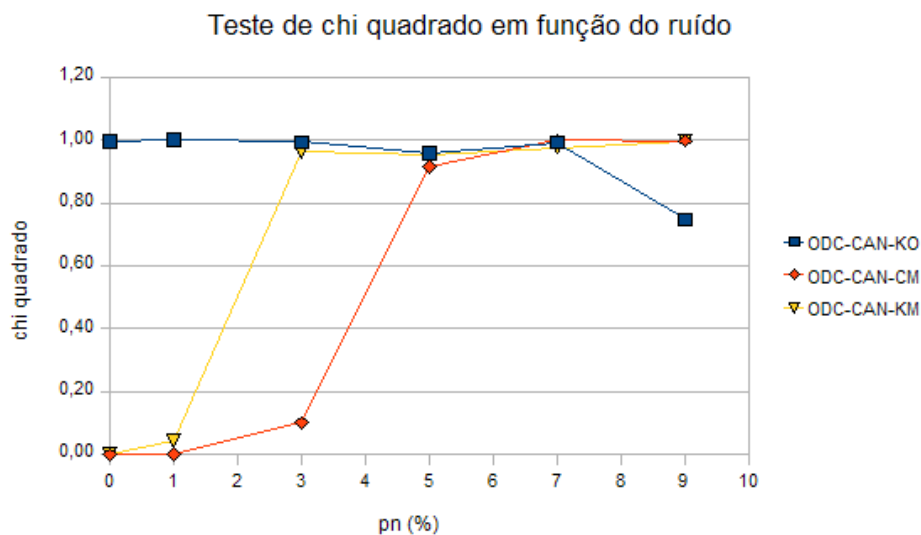


Figura 13: Graus de similaridade do método ODC-CAN em relação aos métodos KO, CM e KM, usando o teste de  $\chi^2$ , em função do ruído percentual

- [7] W. P. Santos, F. M. Assis, R. E. Souza and P. B. Santos Filho. “Evaluation of Alzheimer’s Disease by Analysis of MR Images using Objective Dialectical Classifiers as an Alternative to ADC Maps”. In *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5506–5509, Vancouver, Canada, 2008. EMBS-IEEE.
- [8] S. Haykin. *Redes Neurais: Princípios e Prática*. Bookman, Porto Alegre, 2001.
- [9] W. P. Santos, R. E. Souza, A. F. D. Silva and P. B. Santos Filho. “Evaluation of Alzheimer’s disease by analysis of MR images using multilayer perceptrons, polynomial nets and Kohonen LVQ classifiers”. In *Lecture Notes in Computer Science: Computer Vision / Computer Graphics Collaboration Techniques and Applications (MIRAGE 2007)*, volume 2, pp. 12–22, Rocquencourt, France, 2007. INRIA & CS-IEEE.
- [10] W. P. Santos, R. E. Souza, A. F. D. Silva and P. B. Santos Filho. “Evaluation of Alzheimer’s disease by analysis of MR images using multilayer perceptrons, polynomial nets and Kohonen LVQ classifiers”. *Lecture Notes in Computer Science*, vol. 2, no. 4418, pp. 12–22, 2007.
- [11] W. P. Santos, R. E. Souza, A. F. D. Silva and P. B. Santos Filho. “Avaliação da Doença de Alzheimer pela Análise Multiespectral de Imagens DW-MR por Mapas Auto-Organizados de Kohonen como Alternativa aos Mapas ADC”. In *IV Congresso Latino-Americano de Engenharia Biomédica*, pp. 399–402, Porlamar, Venezuela, 2007. IFMBE.

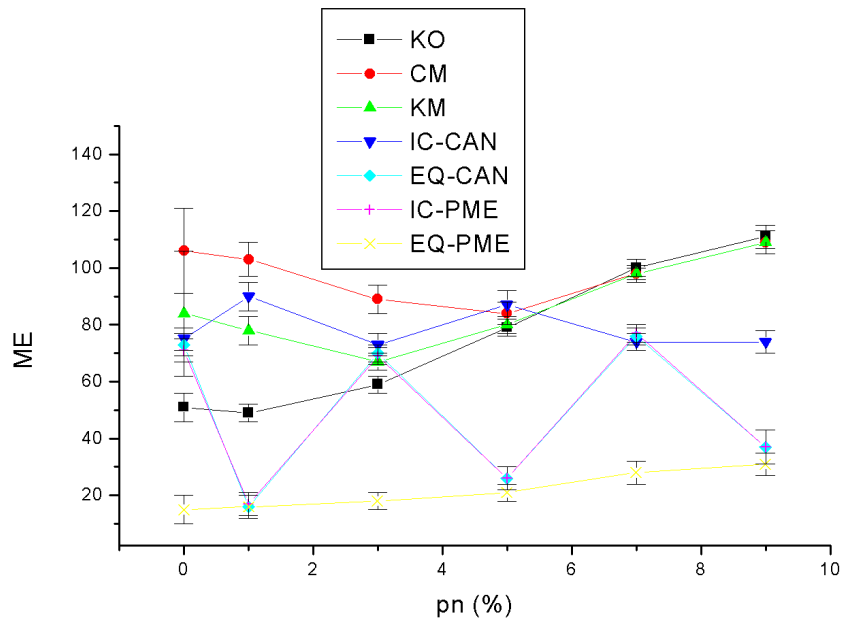


Figura 14: Resultados de  $\epsilon_{ME}$  em função do ruído percentual (ODM)

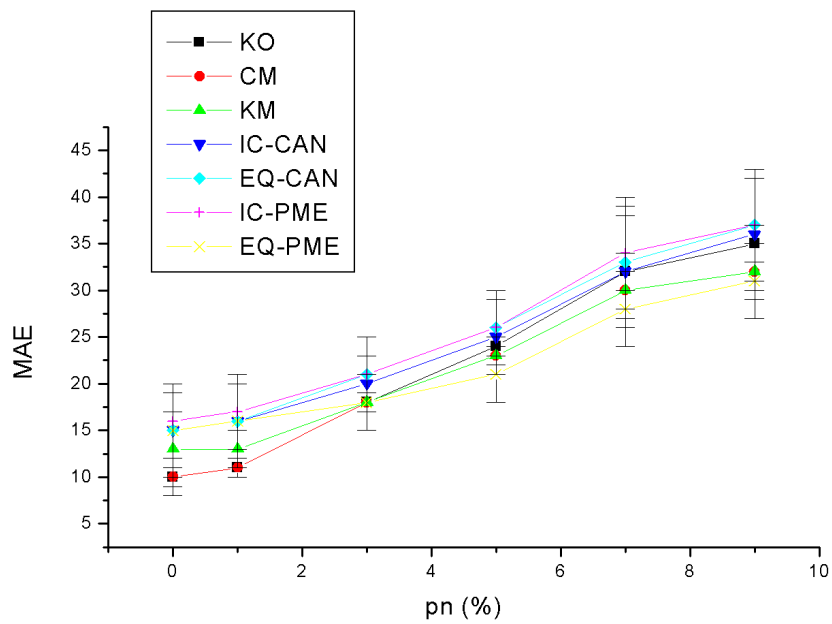


Figura 15: Resultados de  $\epsilon_{MAE}$  em função do ruído percentual (ODM)

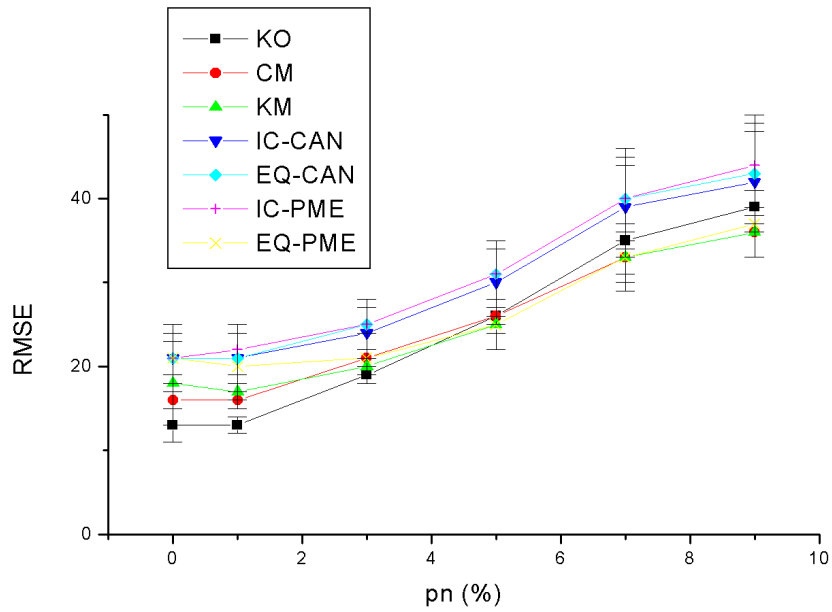


Figura 16: Resultados de  $\epsilon_{RMSE}$  em função do ruído percentual (ODM)

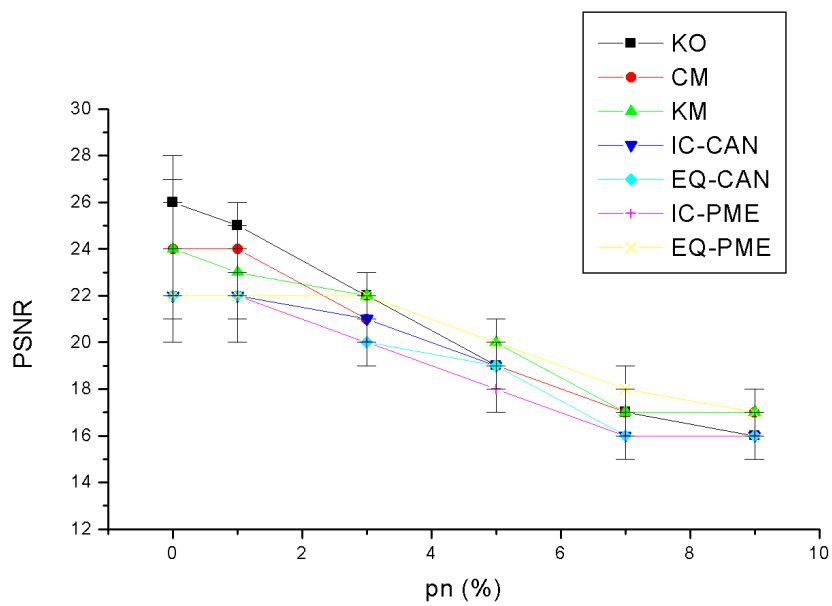


Figura 17: Resultados de  $\epsilon_{PSNR}$  em função do ruído percentual (ODM)

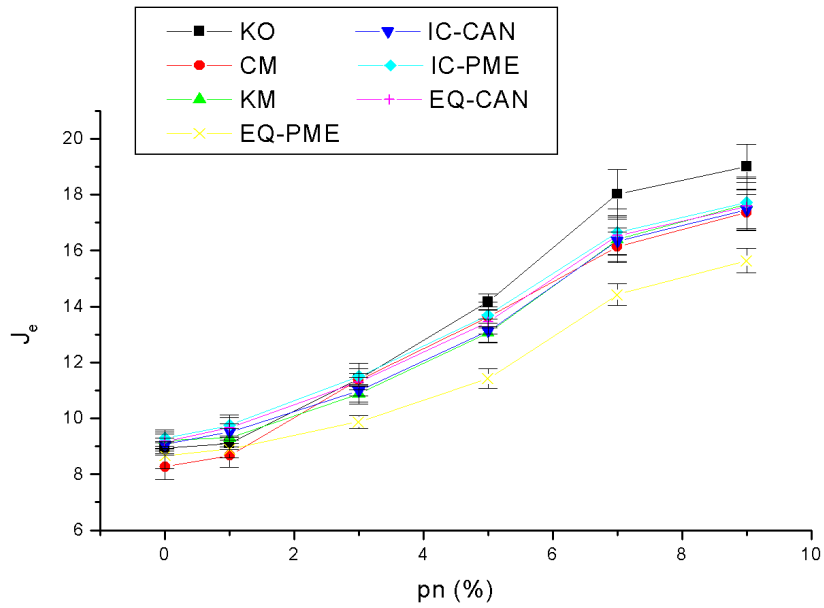


Figura 18: Resultados de  $J_e$  em função do ruído percentual

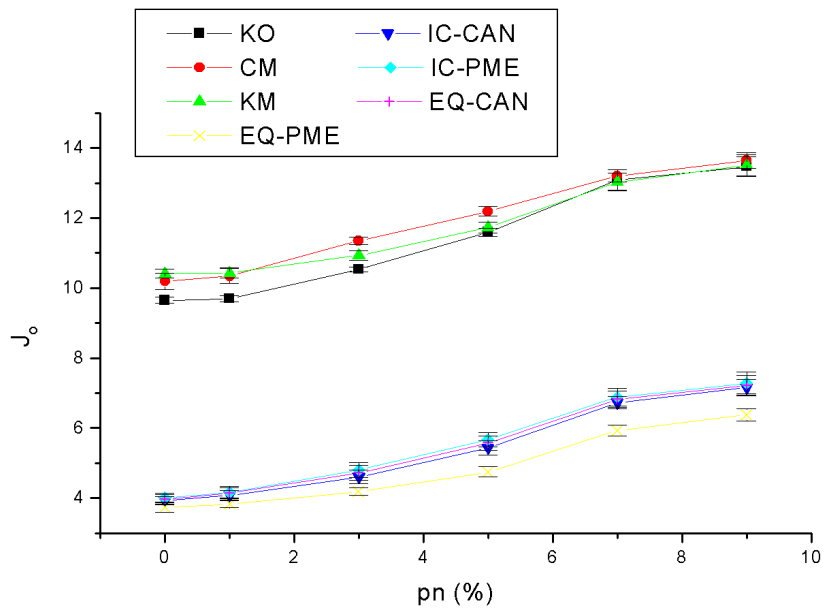


Figura 19: Resultados de  $J_o$  em função do ruído percentual

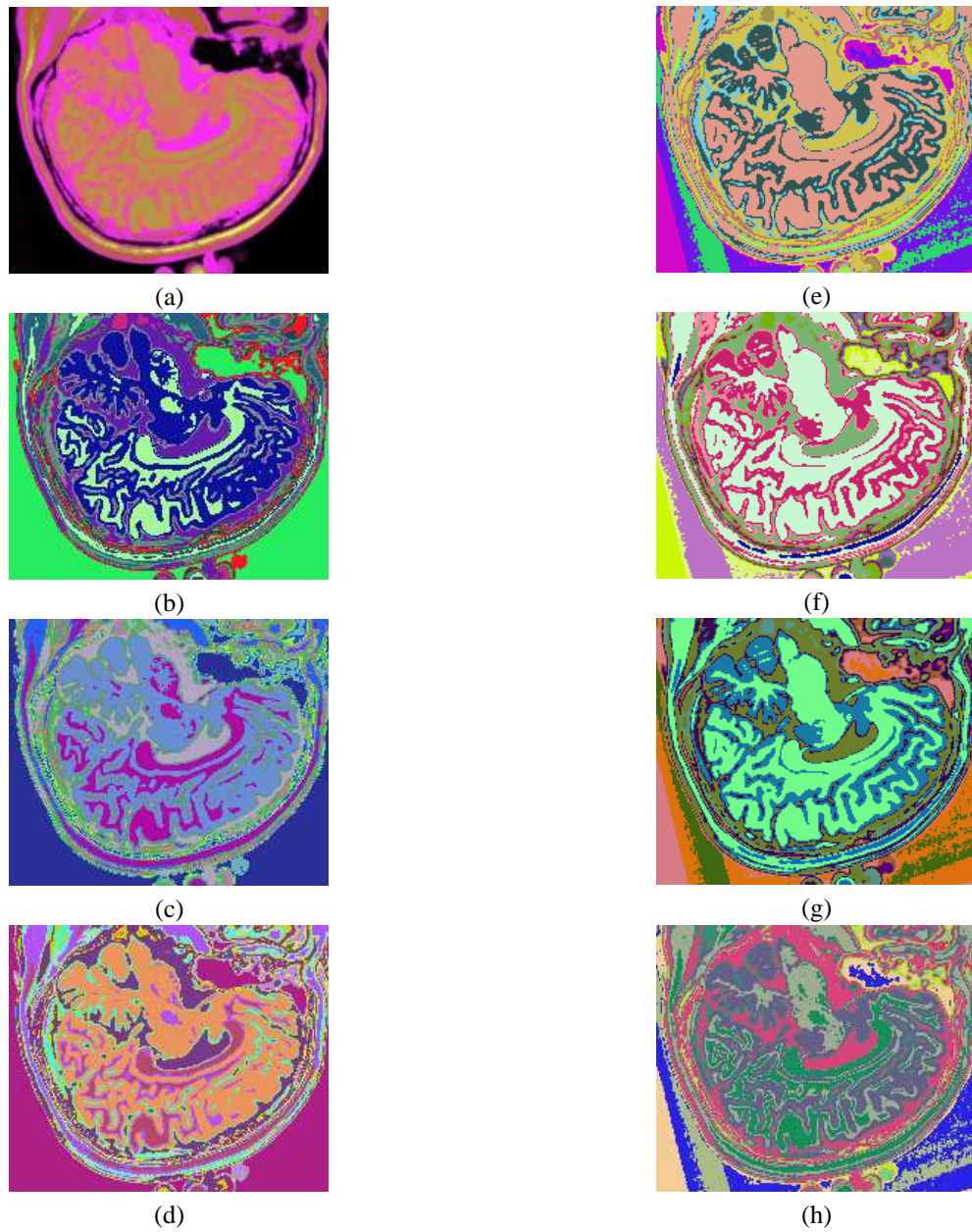


Figura 20: Composição colorida R0-G1-B2 das imagens da fatia 97 ponderadas em PD,  $T_1$  e  $T_2$  (a) e resultados de classificação usando os métodos KO (b), CM (c), KM (d), IC-CAN-KM (e), IC-PME-KM (f), EQ-CAN-KM (g) e EQ-PME-KM (h)



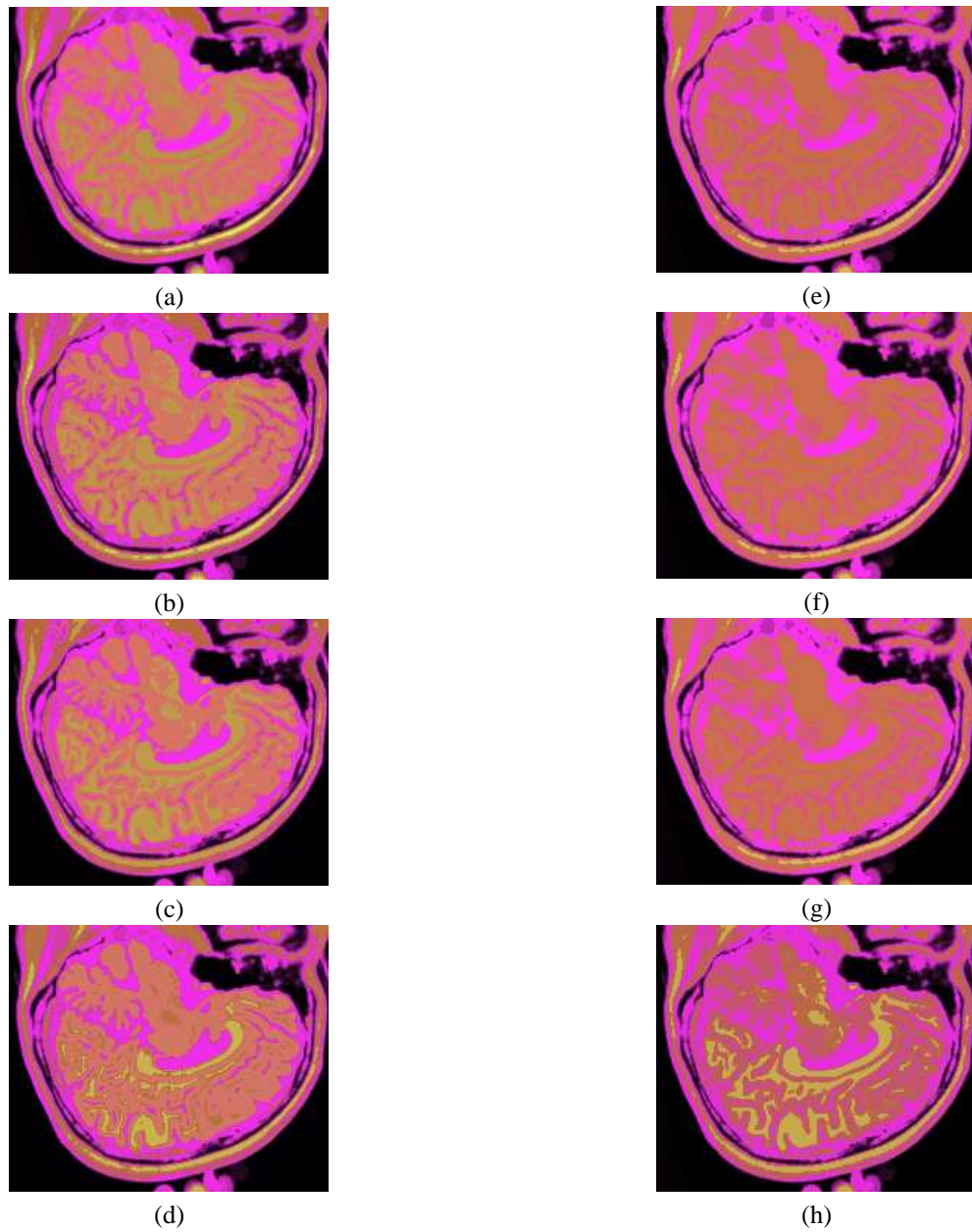


Figura 21: Composição colorida R0-G1-B2 das imagens da fatia 97 ponderadas em PD,  $T_1$  e  $T_2$  (a) e resultados de quantização usando os métodos KO (b), CM (c), KM (d), IC-CAN-KM (e), IC-PME-KM (f), EQ-CAN-KM (g) e EQ-PME-KM (h)

- [12] W. P. Santos, R. E. Souza, A. F. D. Silva and P. B. Santos Filho. “Avaliação da doença de Alzheimer pela análise multi-espectral de imagens DW-MR por redes RBF como alternativa aos mapas ADC”. In *VIII Congresso Brasileiro de Redes Neurais*, Florianópolis, Brasil, 2007. SBRN.
- [13] W. P. Santos, R. E. Souza and P. B. Santos Filho. “Evaluation of Alzheimer’s disease by analysis of MR images using multilayer perceptrons and Kohonen SOM classifiers as an alternative to the ADC maps”. In *29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2118–2121, Lyon, France, 2007. EMBS-IEEE.
- [14] W. P. Santos, R. E. Souza, A. F. D. Silva, N. M. Portela and P. B. Santos Filho. “Análise multiespectral de imagens cerebrais de ressonância magnética ponderadas em difusão usando lógica nebulosa e redes neurais para avaliação de danos causados pela doença de Alzheimer”. In *XI Congresso Brasileiro de Física Médica*, Ribeirão Preto, Brasil, 2006. Sociedade Brasileira de Física Médica.
- [15] A. M. Dan Pelleg. “X-means: Extending K-means with Efficient Estimation of the Number of Clusters”. In *Proceedings of the Seventeenth International Conference on Machine Learning*, pp. 727–734, San Francisco, 2000. Morgan Kaufmann.
- [16] W. P. Santos and F. M. Assis. “Optimization based on Dialectics”. In *IEEE International Joint Conference on Neural Networks (IJCNN 2009)*, pp. 2804–2811, Atlanta, 2009. CIS-IEEE.
- [17] W. P. Santos and F. M. Assis. “Método Dialético de Otimização Usando o Princípio de Máxima Entropia”. *Learning and Nonlinear Models*, vol. 7, pp. 54–64, 2009.
- [18] W. P. Santos and F. M. Assis. “Método Dialético de Otimização Usando o Princípio da Máxima Entropia”. In *IX Congresso Brasileiro de Redes Neurais e Inteligência Computacional*, Ouro Preto, Brazil, 2010.
- [19] W. P. Santos, F. M. Assis, R. E. Souza, P. B. Mendes, H. S. S. Monteiro and H. D. Alves. “Fuzzy-Based Dialectical Non-Supervised Image Classification and Clustering”. *International Journal of Hybrid Intelligent Systems*, vol. 7, pp. 115–124, 2010.
- [20] W. P. Santos, F. M. Assis and R. E. Souza. “MRI Segmentation using Dialectical Optimization”. In *31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5752–5755, Minneapolis, USA, 2009. EMBS-IEEE.
- [21] A. S. Vázquez. *Filosofia da Práxis*. Expressão Popular, São Paulo, 2007.
- [22] L. Konder. *O que é Dialética*. Brasiliense, São Paulo, 2004.
- [23] S. Rahnamayan, H. R. Tizhoosh and M. M. A. Salama. “A novel population initialization method for accelerating evolutionary algorithms”. *Computers and Mathematics with Applications*, , no. 53, pp. 1605–1614, 2007.
- [24] W. L. Hung, M. S. Yang and D. H. Chen. “Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation”. *Pattern Recognition Letters*, , no. 27, pp. 424–438, 2006.
- [25] W. Chen, M. L. Giger and U. Bick. “A Fuzzy C-Means (FCM)-Based Approach for Computerized Segmentation of Breast Lesions in Dynamic Contrast-Enhanced MR Images”. *Academic Radiology*, vol. 13, no. 1, pp. 63–72, 2006.
- [26] C. Windischberger, M. Barth, C. Lamm, L. Schroeder, H. Bauer, R. C. Gur and E. Moser. “Fuzzy cluster analysis of high-field functional MRI data”. *Artificial Intelligence in Medicine*, , no. 29, pp. 203–223, 2003.
- [27] M. D. Alexiuk and N. J. Pizzi. “Robust centroids using fuzzy clustering with feature partitions”. *Pattern Recognition Letters*, , no. 26, pp. 1039–1046, 2005.
- [28] C. Zhu and T. Jiang. “Multicontext fuzzy clustering for separation of brain tissues in magnetic resonance images”. *NeuroImage*, , no. 18, pp. 685–696, 2003.
- [29] M. Jiang, Y. P. Luo and S. Y. Yang. “Stochastic convergence analysis and parameter selection of the standard particle swarm optimization algorithm”. *Information Processing Letters*, vol. 102, pp. 8–16, 2007.
- [30] Y. L. Zheng, L. H. Ma, L. Y. Zhang and J. X. Qian. “On the convergence analysis and parameter selection in particle swarm optimization”. In *Proceedings of the Second International Conference on Machine Learning and Cybernetics*, pp. 1802–1807, Xian, China, 2003.
- [31] I. C. Trelea. “The particle swarm optimization algorithm: convergence analysis and parameter selection”. *Information Processing Letters*, , no. 85, pp. 317–325, 2003.
- [32] L. F. Q. Silveira, L. G. Q. Silveira-Júnior, F. M. Assis and E. L. Pinto. “Algoritmo genético aplicado na obtenção de constelações de sinais para transmissão digital com codificação wavelet em canais com desvanecimento Rayleigh”. In *VIII Congresso Brasileiro de Redes Neurais*, Florianópolis, Brasil, 2007.

- [33] R. Poli and W. B. Langdon. “Markov chain models of bare-bones particle swarm optimizers”. In *Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation*, pp. 142–149, London, 2007.
- [34] E. F. Campana, G. Fasano and A. Pinto. “Dynamic system analysis and initial particles position in particle swarm optimization”. In *IEEE Swarm Intelligence Symposium*, Indianapolis, USA, 2006.
- [35] C. J. F. T. Braak. “A Markov Chain Monte Carlo version of the genetic algorithm Differential Evolution: easy Bayesian computing for real parameter spaces”. *Statistics and Computing*, vol. 16, no. 3, pp. 239–249, 2006.
- [36] E. Cantú-Paz. “Markov chain models of parallel genetic algorithms”. *IEEE Transactions on Evolutionary Computation*, vol. 4, no. 3, pp. 216–226, 2000.
- [37] A. H. Wright and Y. Zhao. “Markov chain models of genetic algorithms”. In *Proceedings of the Genetic and Evolutionary Computation (GECCO) Conference*, pp. 734–741. Morgan Kaufmann, 1999.
- [38] J. Suzuki. “A further result on the Markov chain model of genetic algorithms and its application to a simulated annealing-like strategy”. *IEEE Transactions Systems, Man, Cybernetics*, vol. 28, pp. 95–102, 1998.
- [39] G. Rudolph. “Convergence Analysis of Canonical Genetic Algorithms”. *IEEE Transactions on Neural Networks*, vol. 5, no. 1, pp. 96–101, 1994.
- [40] W. P. Santos, F. M. Assis, P. B. Mendes, H. S. S. Monteiro and H. D. Alves. “An Object-Oriented Approach to Design Dialectical Systems”. In *Horizons in Computer Science Research*, edited by T. S. Clary, New York, 2009. Nova Science.
- [41] D. Chen and C. Zhao. “Data-driven fuzzy clustering based on maximum entropy principle and PSO”. *Expert Systems with Applications*, vol. 36, pp. 625–633, 2009.
- [42] K. Rose. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2210–2239, 1998.
- [43] K. Rose, E. Gurewitz and G. C. Fox. “Statistical mechanics and phase transitions in clustering”. *Physical Reviews Letters*, vol. 65, no. 8, pp. 945–948, 1990.
- [44] R. Zhang and J. Liu. “Underwater image segmentation with maximum entropy based on particle swarm optimization (PSO)”. In *Proceedings of the First International Multi-Symposium on Computer and Computational Sciences (IM-SCCS’06)*, 2006.
- [45] D. H. Wolpert and W. G. Macready. “No free lunch theorems for optimization”. *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [46] M. Omran, A. P. Engelbrecht and A. Salman. “Particle swarm optimization method for image clustering”. *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 19, no. 3, pp. 297–321, 2005.
- [47] D. W. van der Merwe and A. P. Engelbrecht. “Data clustering using particle swarm optimization”. In *IEEE Congress on Evolutionary Computation*, pp. 185–191, Canberra, Australia, 2003. CIS-IEEE.
- [48] H. Pedrini and W. R. Schwartz. *Análise de Imagens Digitais: Princípios, Algoritmos e Aplicações*. Thomson, São Paulo, 2008.
- [49] R. Duda, P. Hart and D. G. Stork. *Pattern Classification*. John Wiley and Sons, 2001.
- [50] R. Duda and P. Hart. *Pattern Classification and Scene Analysis*. John Wiley and Sons, 1972.
- [51] R. K. S. Kwan, A. C. Evans and G. B. Pike. “MRI simulation-based evaluation of image-processing and classification methods”. *IEEE Transactions on Medical Imaging*, vol. 18, no. 11, pp. 1085–1097, 1999.
- [52] D. L. Collins, A. P. Zijdenbos, V. Kollokian, J. G. Sled, N. J. Kabani, C. J. Holmes and A. C. Evans. “Design and Construction of a Realistic Digital Brain Phantom”. *IEEE Transactions on Medical Imaging*, vol. 17, no. 3, pp. 463–468, 1998.
- [53] S. Das, A. Abraham and A. Konar. “Automatic clustering using an improved differential evolution algorithm”. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, vol. 38, no. 1, pp. 218–237, 2008.
- [54] S. Das, S. Sil and U. K. Chakraborty. “Kernel-Based Clustering of Image Pixels with Modified Differential Evolution”. In *IEEE World Congress on Computational Intelligence (WCCI 2008)*, pp. 3472–3479, Hong Kong, China, 2008. CIS-IEEE.
- [55] S. Das and A. Konar. “Automatic image pixel clustering with an improved differential evolution”. *Applied Soft Computing*, no. 9, pp. 226–236, 2009.

- [56] Z. Wang and A. C. Bovik. “A universal image quality index”. *IEEE Signal Processing Letters*, vol. 9, 2002.
- [57] S. Saha and S. Bandyopadhyay. “Application of a New Symmetry-Based Cluster Validity Index for Satellite Image Segmentation”. *IEEE Geoscience and Remote Sensing Letters*, vol. 5, no. 2, pp. 166–170, 2008.