

Avaliação de qualidade visual baseada em rede Bayesiana aplicada a imagens médicas.

Luana Gonçalves
Aldebaro B. R. Klautau Júnior
Ronaldo de Freitas Zampolo
Instituto de Tecnologia
Universidade Federal do Pará
Email: luanagoncalves4@live.com

Hugo Abreu Mendes
Escola Politécnica de Pernambuco
Universidade de Pernambuco
Email: ham@poli.br

Gustavo Maximiliano Alves
Instituto de Ciências da Saúde
Universidade Federal do Pará
Email: gustavo.g.ma@hotmail.com

Resumo—Considerando o volume de dados produzidos diariamente por centros medico-hospitalares, algoritmos de compressão para imagens médicas representam uma alternativa para reduzir custos de transmissão e armazenamento em sistemas PACS. Nesse caso, esquemas adaptativos podem maximizar a compressão evitando a perda de qualidade visual e, por conseguinte, o valor diagnóstico das imagens. Entretanto, métricas de qualidade visual não são usualmente desenvolvidas para esse contexto específico. Nesse trabalho, medidas de qualidade baseadas em rede Bayesiana são propostas e validadas para o caso específico de imagens de tomografia de crânio codificadas por JPEG e JPEG 2000. O protocolo de coleta de dados e validação segue as recomendações do ITU.

Palavras-chave—Qualidade visual, imagens médicas, compressão de imagens.

I. INTRODUÇÃO

O desenvolvimento de novas tecnologias tem facilitado o aperfeiçoamento de métodos de aquisição de imagens, assim como, o surgimento de novas modalidades de sinais visuais para fins diagnósticos [1]. A eclosão de tais tecnologias levanta questões importantes a respeito da aquisição, armazenamento, transferência, compressão e exibição de imagens médicas [2]. Entre elas, a qualidade visual da imagem, a qual pode sofrer substanciais perdas durante tais processos. Logo, avaliar a qualidade da imagem precisamente tem se tornado um tópico bastante discutido nos últimos anos [3]–[6]. Por dia, as unidades medico-hospitalares produzem um conjunto significativo de imagens. Gerir toda essa crescente informação gerada, é um desafio para os sistemas de arquivamento de imagens. Além disso, no caso de aplicações de telemedicina, a transmissão de imagens médicas requer infraestrutura de comunicação muitas vezes ausente - em zonas rurais ou afastadas de grandes centros, por exemplo - e grande demanda dado o número reduzido de especialistas disponíveis [7].

Nesse cenário, o ACR (*American College of Radiology*) estabelece que esquemas de compressão irreversível podem ser utilizados para reduzir o volume de dados de imagens de diagnóstico, desde que sob a supervisão de um médico especialista, uma vez que a taxa de compressão tolerada depende do tecido analisado, da modalidade do exame e principalmente do objetivo do estudo [8]. Além disso, o NEMA (*National Electrical Manufacturers Association*) determina que se esquemas

de compressão reversíveis ou irreversíveis forem adotados, somente algoritmos como Joint Photographic Experts Group devam ser utilizados, entre eles JPEG, JPEG-LS, JPEG-2000, ou MPEG. Entretanto, nem o ACR/NEMA, FDA/U.S (*Food and Drug Administration*) ou a documentação do padrão DICOM (*Digital Imaging Communications in Medicine*) estabelecem recomendação, aprovação ou declaração geral sobre as taxas de compressão apropriadas para qualquer modalidade, doença ou aplicação clínica específica que assegure o objetivo aceitável para o diagnóstico [9], [10].

Apesar da maneira mais confiável de avaliação de qualidade clínica ainda ser baseada em experimentos subjetivos com médicos especialistas, essa tarefa é pouco prática e dispendiosa. Além do mais, as respostas podem variar entre profissionais. Nesse contexto, algoritmos de métricas de qualidade visual (VQM, *visual quality metric*) representam alternativa por serem ágeis além da possibilidade de aclopar em sistemas embarcados [11].

A aplicação do conhecimento decorrente de tais estudos na caracterização da qualidade de imagens com objetivo diagnóstico ainda é bastante incipiente, havendo poucos trabalhos abordando esse assunto, dentre os quais destacamos [2], [12]. Como métricas que atuam diretamente na intensidade do pixel da imagem não são capazes de caracterizar bem a qualidade perceptual e modelos baseados no sistema visual humano são limitados em imagens borradas, a caracterização de algoritmos baseados nas informações de gradiente da imagem vem crescendo para imagens médicas [2].

Em [13]–[15], métricas baseadas na direção ou magnitude do gradiente da imagem apresentam boa correlação com avaliações subjetivas. Em contrapartida, propostas baseadas na associação da magnitude e direção do gradiente da imagem em um único coeficiente são apresentadas em [16], [17], onde o foco é analisar a sensibilidade às mudanças de detalhes nas bordas das imagens. A comunidade radiológica, entretanto, ainda não adotou um método padrão ou protocolo de avaliação objetiva para aferição de qualidade clínica.

O presente trabalho apresenta uma nova métrica de qualidade visual específica para imagens médicas codificadas a partir do desenho de uma rede Bayesiana, dada a explícita natureza estocástica das avaliações subjetivas. Tal métrica pode

constituir instrumento de caracterização de qualidade por meio dos quais algoritmos e sistemas de aquisição e processamento podem ser otimizados.

II. SISTEMA PARA COLETA DE DADOS

O sistema proposto para coleta de dados foi implementado levando em consideração as etapas pelas quais as imagens passam, desde a aquisição até a emissão do laudo diagnóstico, segundo as recomendações do ITU (*International Telecommunication Union*) [18] e os critérios definidos no *European Guidelines on Quality Criteria for Computed Tomography* [10]. Imediatamente após a aquisição, um técnico ou médico analisa determinados parâmetros da imagem, tais como o grau de precisão dos limites entre estruturas que a compõem, ou o quão regular se mostram seus contornos. A análise desses aspectos teria por finalidade indicar se a imagem apresenta artefatos, ruídos ou distorções. Nesta etapa, o foco é determinar se a imagem em questão apresenta qualidade suficiente para permitir um diagnóstico adequado. Em caso positivo, passa-se então à etapa de laudo, onde serão avaliadas as estruturas anatômicas imageadas. Caso contrário, a fase de aquisição é repetida. Considerando que o laudo é emitido no mesmo momento que a aquisição da imagem, a compressão da imagem é desnecessária. Se o exame, entretanto, for transmitido ou armazenado técnicas de compressão podem ser usadas, desde que não haja perda de qualidade clínica.

Em [10], o *European Study Group of Radiologists and Physicists* estabelece quais os critérios de qualidade de exames da imagem que garantem a preservação de toda informação clínica necessária ao diagnóstico. Tais critérios são utilizados por médicos ou técnicos radiologistas nas rotinas de exames para definir se a imagem possui o padrão de qualidade de diagnóstico. Como a qualidade difere para cada técnica é necessário que hajam diferentes critérios para cada tipo de exame, isso porque algumas áreas anatômicas em determinados exames não interferem no diagnóstico e podem ser visualizadas com baixa qualidade em relação a outras. Na Tabela I, são listados os critérios de reprodução visual definidos para imagens de tomografia de crânio.

O conjunto de imagens para avaliação é obtido a partir de 3 imagens originais de tomografia computadorizada (CT) de crânio, com cortes no plano transversal, dentre as quais 1 imagem sem indicação de estruturas patológicas e 2 imagens com patologias inerentes, todas em formato DICOM (dcm). Para cada imagem original, foram produzidas duas outras versões, mediante janelamento padrão para osso e tecidos moles (Fig. 1). A escolha de janelas para osso e tecidos moles é justificada por serem as opções padrão em análise de imagens de CT de crânio. As imagens originais foram adquiridas nas dimensões em 512×512 e 16-bits de profundidade.

As imagens foram codificadas utilizando os algoritmos JPEG (fator de qualidade = 50, 60, 70 e 100) e JPEG2000 (fator de qualidade = 30, 40, 50 e 100), em diferente níveis de compressão (Fig. 2).

Ferramentas de processamento de imagens foram utilizadas para modificar cada slice com as características em que os

Tabela I

CRITÉRIOS DE REPRODUÇÃO VISUAL DE QUALIDADE PARA EXAMES DE TOMOGRAFIA COMPUTADORIZADA DE CRÂNIO

Critérios para tecidos moles	
Reprodução crítica dos limite entre a massa branca e a cinzenta	
Reprodução crítica dos gânglios da base	
Reprodução crítica do sistema ventricular	
Reprodução crítica do espaço do líquido cefalorraquidiano em volta do mesencéfalo	
Reprodução crítica do espaço do líquido cefalorraquidiano em volta do cérebro	
Reprodução crítica dos grandes vasos e plexo coroide após de contraste intravenoso	
Critérios para osso	
Reprodução crítica das estruturas corticais e trabeculares dos ossos	
Reprodução crítica dos compartimentos gasosos	
Reprodução crítica da sela túrcica	
Reprodução crítica dos contornos cerebelares	
Reprodução crítica dos limites entre substância branca e cinzenta do cerebelo	
Reprodução do espaço do líquido cefalorraquidiano em volta do tronco encefálico	
Reprodução crítica dos grandes vasos e plexo coroide após de contraste intravenoso	

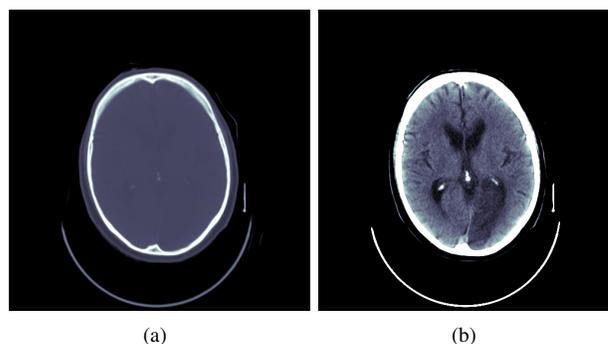


Figura 1. Slice com diferentes janelamentos: (a) osso; (b) tecidos moles.

laudos médicos são emitidos em rotinas clínicas. O resultado foram imagens com 1024×1024 e 8-bits de profundidade, ampliadas via interpolação bicúbica e codificadas em ambos esquemas de compressão (JPEG e JPEG2000). Em cada rodada de avaliação são apresentadas 48 imagens, as quais não podem ser alteradas. O conjunto de avaliadores consiste em dois especialistas em diagnóstico por imagens de crânio.

O experimento foi implementado em linguagem Python, e o aplicativo de código aberto PsychoPy [19] e pode ser dividido nas etapas ilustradas na Fig. 3.

Inicialmente são apresentadas orientações a respeito do procedimento e seu objetivo. Em seguida, uma fase de aquecimento familiariza o avaliador com a interface do sistema. Posteriormente, uma notificação é exibida indicando o fim do aquecimento. E por último, a fase de coleta de dados. Nesta, as imagens que compõem o conjunto de teste são apresentadas aleatoriamente, uma de cada vez. Para cada imagem exibida, o avaliador deve responder à pergunta: “Qual o grau de visibilidade das estruturas anatômicas da imagem apresentada?”, e pontuar a qualidade da imagem de 1 a 5, cada critério listado na Tabela I de acordo com a escala listada na Tabela II [20].

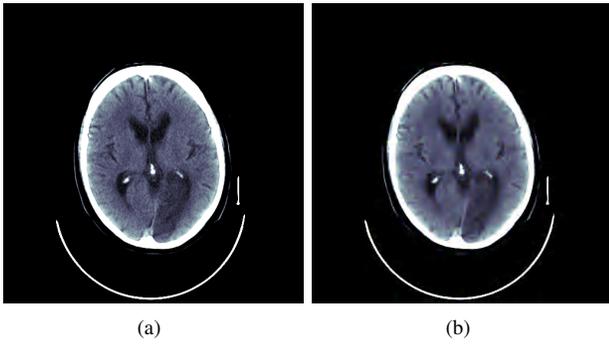


Figura 2. Exemplo de imagem codificada: (a) imagem original; (b) imagem reconstruídas.

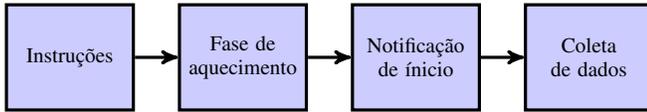


Figura 3. Etapas do procedimento experimental.

Para aquisição das imagens, dados de avaliação e continuidade da pesquisa, o projeto intitulado “Avaliação de qualidade visual aplicada a imagens de tomografia computadorizada” foi submetido ao Comitê de ética em pesquisa do Instituto de Ciências da Saúde da Universidade Federal do Pará - ICS/ UFPA, o qual emitiu um parecer substanciado definindo que o protocolo encaminhado dispõe de metodologia e critérios definidos conforme a resolução 466/12 do Conselho Nacional de Saúde do Ministério da Saúde e, portanto, aprovando a continuidade do projeto e consequentemente a coleta de dados ¹. O projeto é mantido em repositório online², onde os interessados poderão obter a versão estável mais recente, contribuir com sugestões ou mesmo colaborar no desenvolvimento do aplicativo (Fig. 4).

A qualidade final da imagem (OS, *Opinion Score*) é dada por:

$$OS = \frac{\sum_{s=1}^S \sum_{t=1}^T G_{abs,s,t}}{S \cdot T}, \quad (1)$$

¹CAAE: 86587118.4.0000.0018.

²<https://github.com/luaffjk/AQVIM>

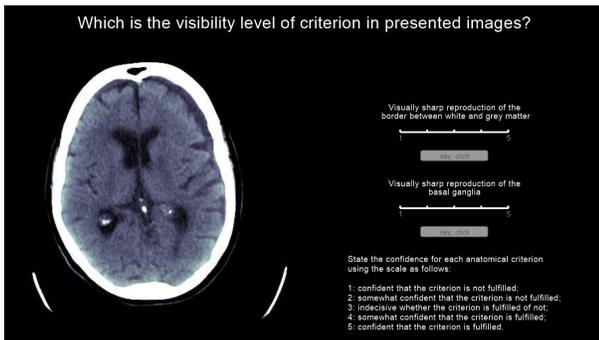


Figura 4. Interface gráfica de usuário para coleta de dados de imagens.

Tabela II
ESCALA PARA AVALIAÇÃO DE IMAGENS CLÍNICAS

Escala	Aparência das Estruturas
1	Qualidade certamente não compromete o diagnóstico
2	Qualidade possivelmente não compromete o diagnóstico
3	Qualidade inconclusiva
4	Qualidade possivelmente compromete o diagnóstico
5	Qualidade certamente compromete o diagnóstico

onde $G_{abs,s,t}$ é a taxa (Tab. II) em relação à aparência de uma estrutura em particular (Tab. I) dada por S e T representa o número de avaliadores [20]. O range dinâmico de $OS \in [1, 5]$.

III. MEDIDA DE QUALIDADE BAYESIANA

A percepção visual da imagem não é um processo único, e sim uma tarefa multifacetada, que envolve a combinação de características que engendram a qualidade global da estrutura imageada. Porém, a síntese dessa combinação do ponto de vista perceptual, cognitivo e psicológico não é meramente uma soma dessas características, o que explica a variância entre avaliações obtidas a partir de diferentes observadores. Logo, como as propriedades mensuráveis da imagem e as observações humanas registradas devem ser conectadas? É possível explicar ou simular o fenômeno da qualidade visual? [21]

Nesse sentido, a combinação de métricas de avaliação de qualidade visual é proposta a partir da estrutura de uma rede Bayesiana, dada a explícita natureza estocástica das avaliações subjetivas. A motivação do uso de redes de decisão é fundamentada, primeiro, porque pode-se verificar que a maioria dos modelos para avaliação de qualidade visual é determinística, muito embora a avaliação subjetiva esteja sujeita a fatores aleatórios, como preferências culturais e pessoais do avaliador [22].

Além disso, no caso da necessidade de atualização se verificado que o modelo escolhido se tornou defasado face a novos dados experimentais, é possível otimizar os parâmetros da função, ou seja o modelo pode aprender com base em novas experiências apresentadas, e se adaptar a novas perspectivas, caracterizando um paradigma de aprendizado de máquina.

A. Redes Bayesianas

Redes Bayesianas são grafos acíclicos orientados capazes de modelar conhecimentos sujeitos à aleatoriedade, como uma das suas principais características a adaptabilidade. O objetivo de uma inferência Bayesiana é produzir um agente mutável que tenha desempenho semelhante às decisões humanas em tarefas de complexidade intelectual, que pode se adaptar a ambientes estocásticos e instáveis, e lidar de maneira sensata com variadas fontes de incertezas [22], [23].

A estrutura de uma rede de Bayes é composta por nós de entrada que são as variáveis causais do evento que será modelado e nós de saída que são dados baseados na decisão de especialistas de casos passados observados. A ideia básica é treinar uma estrutura de árvore de decisão, modelando

uma rede multi-conectada que tem por finalidade retornar a probabilidade de cada possibilidade dos nós de saída [21].

No treinamento de uma rede, a primeira etapa, é identificar as variáveis relevantes que irão compor os nós de entrada da rede, e dispor de uma base de dados fornecidos por especialistas ou proveniente de dados de campo coletados. Em seguida, são estimadas as probabilidades a priori das variáveis de entrada em relação às saídas, bem como as probabilidades das saídas. Então, a rede calcula as probabilidades a posteriori das saídas, dada cada configuração dos seus nós de entrada, através do teorema de Bayes definido por:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B|A_1) + P(B|A_2) + \dots + P(B|A_n)}, \quad (2)$$

onde $P(A)$ é a probabilidade de um evento A , o conjunto $[A_1, A_2, \dots, A_n]$ representa uma partição do evento esperado e B é um evento arbitrário [24]. A classe com maior probabilidade a posteriori é escolhida como saída da rede.

B. Medida de qualidade proposta

Dada a diferença perceptual entre imagens codificadas por algoritmos de compressão distintos, a tarefa foi dividida em relação ao algoritmo de compressão da imagem. Com isso, duas redes foram modeladas: uma para imagens codificadas através de JPEG e outra para imagens com padrão de compressão JPEG 2000.

No modelo proposto, três métricas de qualidade visual servem como entrada do sistema, cuja saída é a qualidade percebida da imagem que se quer avaliar. São elas: *Gradient map weighting-gradient direction similarity* (GMW-GDSIM), *Edge preservation measure* (EPM) e o *Mean structural similarity index* (MSSIM) [16], [17], [25]. Na Figura 5, o fluxograma do desenho do modelo é apresentado.

Na Tabela III, é apresentada a adaptação entre as possíveis saídas das redes de acordo com a qualidade percebida (Tab. II).

Na fase de treinamento, uma vez que as métricas que servem como entrada da rede são calculadas, as probabilidades a priori foram estimadas usando-se frequência relativa. A probabilidade da qualidade também foi estimada a partir da

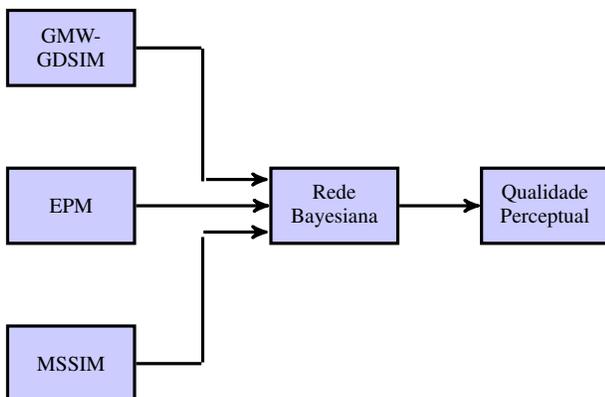


Figura 5. Desenho da métrica proposta para avaliação de qualidade visual de imagens médicas.

Tabela III
SAÍDA DA REDE BAYESIANA EM RELAÇÃO ÀS APARÊNCIAS DAS ESTRUTURAS

Saída da rede	Aparência das Estruturas
Não compromete	Qualidade certamente não compromete diagnóstico
	Qualidade possivelmente não compromete diagnóstico
Inconclusivo	Qualidade inconclusiva
Compromete	Qualidade possivelmente compromete diagnóstico
	Qualidade certamente compromete diagnóstico

frequência relativa. O modelo proposto foi implementado em linguagem Python, através do pacote Scikit-Learn, o qual consiste em uma ferramenta destinada para aprendizado de máquina, além de mineração e análise de dados [26].

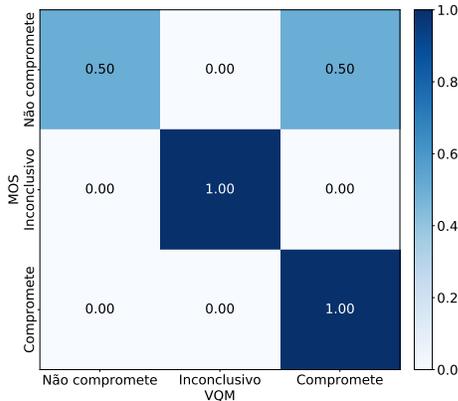
IV. RESULTADOS

O procedimento padrão para caracterização do desempenho de VQM's, consolidado atualmente em documentos do *Video Quality Experts Group* (VQEG) e recomendações do ITU, considera três aspectos: acurácia, monotonicidade e consistência, em termos da relação entre os dados MOS (*mean opinion score*) e a VQM [18], [27]. Para examinar a sensibilidade dos modelos de avaliação de qualidade visual implementados, o comportamento estatístico dos algoritmos foi mapeado e representado na Tabela IV. Como recomendação do relatório do VQEG, a predição de acurácia e monotonicidade são dados através do coeficiente de correlação de Spearman e coeficiente de correlação de Pearson, respectivamente. Adicionalmente são apresentadas a análise de variância (ANOVA) e a acurácia entre os valores de VQM em relação ao MOS para o conjunto de imagens de teste [28], [29]. Como GMW-GDSIM, EPM e MSSIM apresentam faixas dinâmicas contínuas, não é possível calcular a acurácia em relação ao MOS.

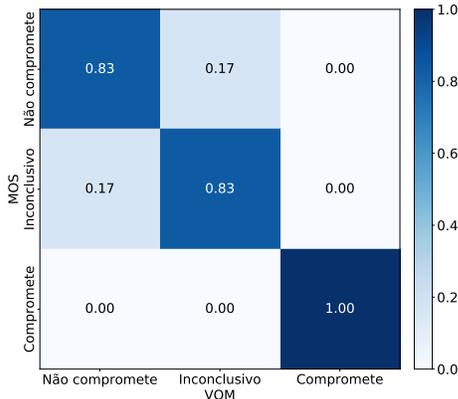
Uma vez que o conjunto de dados esteja desbalanceado, apenas a acurácia e monotonicidade não são suficientes para validar o desempenho de um classificador, isso porque os resultados inferidos podem representar dados tendenciosos. A predição de consistência, por sua vez, relaciona-se com a capacidade da medida de qualidade objetiva fornecer previsões consistentemente precisas para todos os tipos de imagens e não falhar para um subconjunto de imagens. Nesse sentido,

Tabela IV
RESULTADOS DA AVALIAÇÃO DE DESEMPENHO DAS VQM'S EM RELAÇÃO AOS MAPEAMENTOS ESTATÍSTICOS.

Imagens JPEG				
VQM	Spearman	Pearson	ANOVA	Acurácia
GMW-GDSIM	0.14	0.27	35.25	-
EPM	0.32	0.43	31.78	-
MSSIM	0.041	0.20	27.59	-
Rede	0.54	0.51	1.54	0.83
Imagens JPEG 2000				
VQM	Spearman	Pearson	ANOVA	Acurácia
GMW-GDSIM	0.86	0.87	34.02	-
EPM	0.48	0.39	29.73	-
MSSIM	0.49	0.45	25.86	-
Rede	0.91	0.91	1.01	0.88



(a)



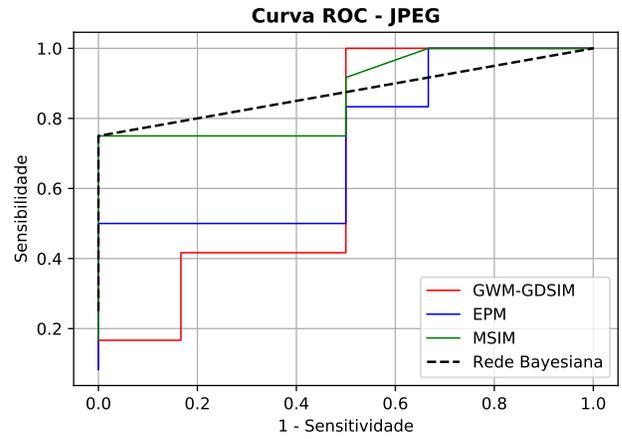
(b)

Figura 6. Matriz de contingência da Rede Bayesiana em relação ao MOS. (a) Imagens comprimidas por JPEG; (b) Imagens comprimidas por JPEG 2000.

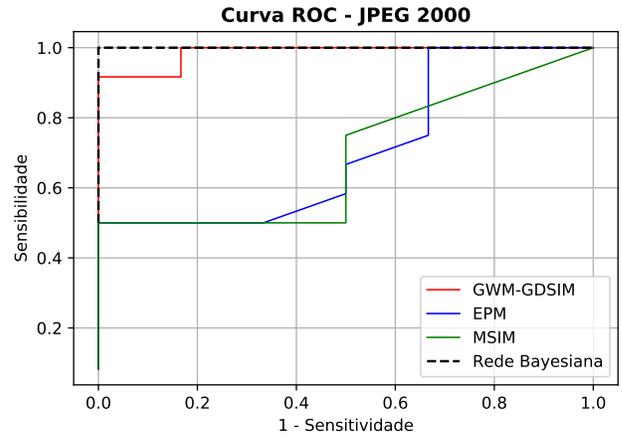
na Figura 6 são apresentadas as matrizes de contingência das saídas das redes Bayesianas em relação ao MOS.

Outra predição de consistência muito utilizada, a curva *Receiver Operating Characteristic* (ROC) é uma ferramenta que estima quão o modelo é capaz de distinguir as classes precisamente, quanto maior a área abaixo da curva (AUC, *area under curve*) melhor é o desempenho do modelo. A técnica relaciona a sensibilidade (*SE*), ou seja, a taxa de amostras que o classificador categoriza como positivas e de fato são positivas, à especificidade (*SP*) dos dados, que consiste na proporção de amostras que o classificador categoriza como negativas que de fato são negativas.

Como a curva ROC só se aplica à classificadores binários, em [30] uma extensão da curva ROC é empregada à tarefas com múltiplas classes, aplicando a abordagem *one vs all*. A técnica consiste em um mapeamento das classes, onde uma delas é definida como positiva e todas as outras classes são unificadas em outra classe única. Na Figura 7 é apresentada a curva ROC das métricas GMW-GDSIM, EPM, MSSIM e Rede Bayesiana em relação ao MOS, utilizando *one vs all*. A classe “Não compromete” continuou definida como “Não compromete”, enquanto que, “Inconclusivo” e “Compromete” foram unificadas para “Compromete”.



(a)



(b)

Figura 7. Curva ROC de VQM's em relação ao MOS. (a) Imagens codificadas por JPEG; (b) Imagens codificadas por JPEG 2000.

Na Figura 8 são apresentadas as matrizes de contingência das redes bayesianas em relação ao MOS utilizando a abordagem *one vs all*.

Em [11], o teste Kolmogorov–Smirnov é utilizado para estimar a performance de VQM's, onde o desempenho de um discriminante é caracterizado pela distância de Kolmogorov–Smirnov (*KS*) dada por:

$$KS(P_1, P_2) = \sup_s |P_1(s) - P_2(s)|, \quad (3)$$

em que $P_1(x)$ é a função de distribuição acumulada do radiologista aceitar a imagem, e $P_2(x)$ representa a função de distribuição acumulada do radiologista não aceitar a imagem, representadas pelas Equações 4 e 5, respectivamente.

$$P_1(s) = 1 - SE \quad (4)$$

$$P_2(s) = 1 - SP \quad (5)$$

Quanto maior a diferença entre as duas distribuições melhor é performance do modelo, entretanto, para calcular a

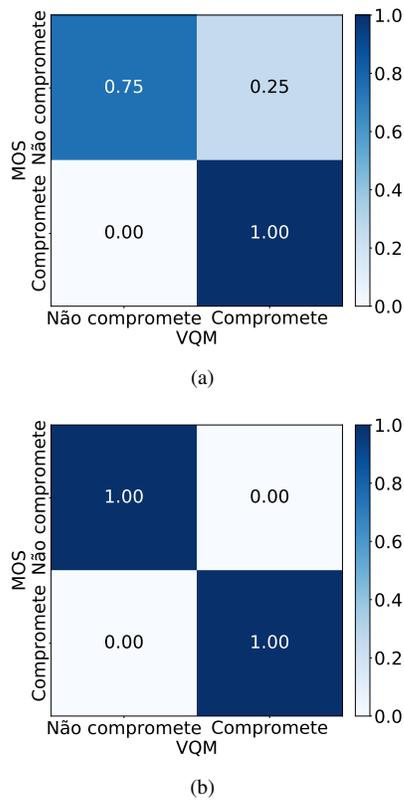


Figura 8. Matriz de contingência da rede Bayesiana em relação ao MOS. (a) Imagens comprimidas por JPEG; (b) Imagens comprimidas por JPEG 2000.

distância de Kolmogorov-Sminov de uma VQM que tenha faixa dinâmica é necessário arbitrar um limiar s' . Nesse caso, s' é escolhido tal que obtenha o máximo KS . Nas Figuras 9 e 10 são apresentados os gráficos do teste de Kolmogorov-Smirnov em relação às VQM's para as imagens comprimidas pelos algoritmos JPEG e JPEG 2000, respectivamente.

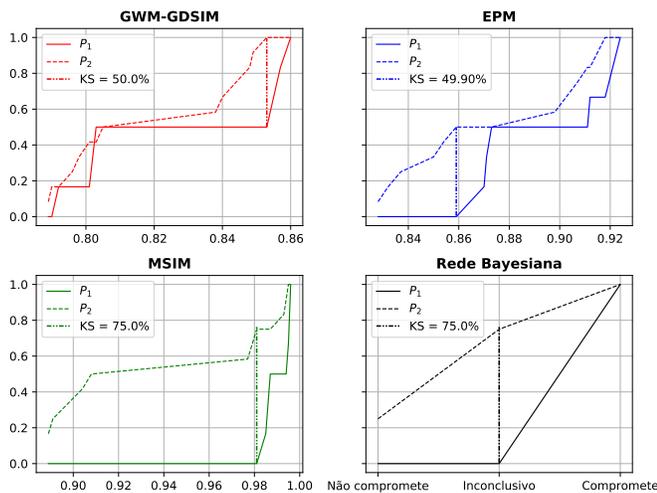


Figura 9. Gráfico do teste de Kolmogorov-Smirnov das VQM's em relação às imagens JPEG.

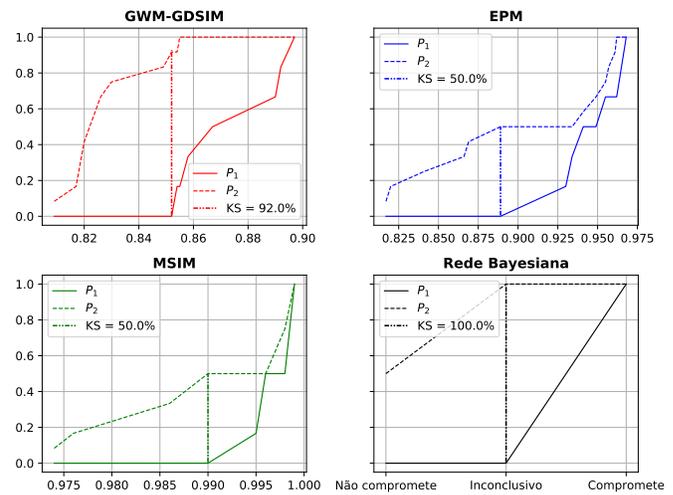


Figura 10. Gráfico do teste de Kolmogorov-Smirnov das VQM's em relação às imagens JPEG 2000.

Na Tabela V a distância KS ; o limiar s' arbitrado e a área abaixo da curva ROC são exibidos para cada VQM em relação ao MOS.

V. CONCLUSÃO

Neste trabalho, foi discutido o problema de avaliação de qualidade visual para imagens médicas codificadas. Foram propostas e validadas experimentalmente medidas de qualidade a partir de modelos Bayesianos. As medidas permitem avaliar a qualidade de imagens de tomografia computadorizada de crânio degradadas por distorções oriundas de algoritmos JPEG e JPEG 2000, possuindo como vantagens manter evidente a natureza estocástica da avaliação subjetiva e ser facilmente atualizada em decorrência de novos dados experimentais.

AGRADECIMENTOS

Os autores agradecem ao CNPq e à UFPA, que financiaram parcialmente este projeto através da concessão de bolsas de pesquisa.

Tabela V
TESTE DE KOLMOGOROV-SMIRNOV.

Imagens JPEG			
VQM	KS	s'	AUC
GMW-GDSIM	50.0%	0.859	0.67
EPM	49.9%	0.853	0.72
MSSIM	75.0%	0.981	0.87
Rede	75.0%	-	0.88
Imagens JPEG 2000			
VQM	KS	s'	AUC
GMW-GDSIM	92.0%	0.852	0.99
EPM	50.0%	0.889	0.71
MSSIM	50.0%	0.99	0.69
Rede	100%	-	1.0

REFERÊNCIAS

- [1] Y. Ding, Ed., *Visual Quality Assessment for Natural and Medical Image*. Springer, 2008.
- [2] C. Cavaro-Menard, L. Zhang, and P. L. Callet, "Diagnostic quality assessment of medical images: Challenges and trends," in *2010 2nd European Workshop on Visual Information Processing (EUVIP)*, July 2010, pp. 277–284.
- [3] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? a new look at signal fidelity measures," *IEEE Signal Processing Magazine*, vol. 26, no. 1, pp. 98–117, Jan 2009.
- [4] L. J. Karam, T. Ebrahimi, S. S. Hemami, T. N. Pappas, R. J. Safranek, Z. Wang, and A. B. Watson, "Introduction to the issue on visual media quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 189–192, April 2009.
- [5] A. Saha and Q. J. Wu, "Full-reference image quality assessment by combining global and local distortion measures," *Signal Processing*, vol. 128, pp. 186 – 197, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0165168416300196>
- [6] H. wen Chang, Q. wen Zhang, Q. gang Wu, and Y. Gan, "Perceptual image quality assessment by independent feature detector," *Neurocomputing*, vol. 151, pp. 1142 – 1152, 2015. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S09252321214013344>
- [7] M. Kaur and V. Wasson, "Roi based medical image compression for telemedicine application," *Procedia Computer Science*, vol. 70, pp. 579 – 585, 2015, proceedings of the 4th International Conference on Eco-friendly Computing and Communication Systems. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1877050915032019>
- [8] L. Zhang, C. Cavaro-Menard, P. L. Callet, and J. Y. Tanguy, "A perceptually relevant channelized joint observer (pcjo) for the detection-localization of parametric signals," *IEEE Transactions on Medical Imaging*, vol. 31, no. 10, pp. 1875–1888, Oct 2012.
- [9] P. Bak, M. Claudon, N. Dugar, D. Harvey, P. H. M. Kämmerer, D. Koff, E. Kotter, T. Kronander, R. Loose, C. Lundström, J. Martin, U.-E. Martin, P. Mildnerberger, D. Müller, J. Reponen, A. Rowberg, L. Sutton, D. Taubman, and A. Tzannes, "Usability of irreversible image compression in radiological imaging. a position paper by the european society of radiology (esr)," *Radiology*, vol. 2, no. 1, p. 103–115, Dec 2011.
- [10] European Commission, "European guidelines on quality criteria for computed tomography," <http://www.dr.dk/guidelines/ct/quality/htmlindex.htm>, 1999, [Online; accessed 08-September-2017].
- [11] I. Kowalik-Urbaniaka, D. Bruneta, J. Wangb, D. Koffc, N. Smolarski-Koffc, E. R. Vrscaya, B. Wallaced, and Z. Wangb, "The quest for "diagnostically lossless" medical image compression: A comparative study of objective quality metrics for compressed medical images," *SPIE Medical Images: Image Perception, Observer Performance, and Technology*, vol. 9037, no. 4, Apr. 2014.
- [12] L. S. Chow and R. Paramesran, "Review of medical image quality assessment," *Biomedical Signal Processing and Control*, vol. 27, pp. 145 – 154, 2016. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1746809416300180>
- [13] Z. Ni, L. Ma, H. Zeng, C. Cai, and K. Ma, "Gradient direction for screen content image quality assessment," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1394–1398, Oct 2016.
- [14] G. Chen, C. Yang, and S. Xie, "Gradient-based structural similarity for image quality assessment," in *2006 International Conference on Image Processing*, Oct 2006, pp. 2929–2932.
- [15] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, Aug 2011.
- [16] W. Preedanana, T. Kondo, P. Bunnun, and I. Kumazawa, "Image quality assessment for medical images based on gradient information," in *2018 5th International Conference on Business and Industrial Research (ICBIR)*, May 2018, pp. 189–194.
- [17] B. P. Bondzulich and V. S. Petrovic, "Edge-based objective evaluation of image quality," in *2011 18th IEEE International Conference on Image Processing*, Sept 2011, pp. 3305–3308.
- [18] ITU-T, *Tutorial – Objective Perceptual Assessment of Video Quality: Full Reference Television*, International Telecommunication Union Geneva, Switzerland, 2004.
- [19] J. W. Peirce, "Psychopy—psychophysics software in Python," *Journal of Neuroscience Methods*, vol. 162, no. 1–2, pp. 8 – 13, 2007. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0165027006005772>
- [20] A. D. Crop, P. T. V. H. Klaus Bacher, P. V. Smeets, B. S. Smet, M. Vergauwen, U. Kiendys, P. Duyck, K. Verstraete, K. D'Herde, and H. Thierens, "Correlation of contrast-detail analysis and clinical image quality assessment in chest radiography with a human cadaver study," *Radiology*, vol. 262, no. 1, pp. 98–304, Jan. 2012.
- [21] T. Eerola, L. Lensu, J.-K. Kamarainen, T. Leisti, R. Ritala, G. Nyman, and H. Kälviäinen, "Bayesian network model of overall print quality: Construction and structural optimisation," *Pattern Recognition Letters*, vol. 32, no. 11, pp. 1558 – 1566, 2011. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0167865511001061>
- [22] R. de Freitas Zampolo and R. Seara, "Perceptual image quality assessment based on bayesian networks," in *Image Processing, 2004. ICIP '04. 2004 International Conference on*, vol. 1, Oct 2004, pp. 329–332 Vol. 1.
- [23] K. B. Korb and A. E. Nicholson, *Bayesian Artificial Intelligence*, 2nd ed. CHAPMAN HALL/CRC, 2003.
- [24] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. McGraw-Hill, 1991.
- [25] B. Kumar, S. B. Kumar, and C. Kumar, "Development of improved ssim quality index for compressed medical images," in *2013 IEEE Second International Conference on Image Information Processing (ICIIP-2013)*, Dec 2013, pp. 251–255.
- [26] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [27] VQEG, "VQEG Final Report of FR-TV Phase II Validation Test," on-line, agosto 2003. [Online]. Available: <http://www.its.bldrdoc.gov/vqeg/projects/frtv-phase-ii/frtv-phase-ii.aspx>
- [28] A. Rencher, *Methods of Multivariate Analysis*, ser. Wiley Series in Probability and Statistics. Wiley, 2003. [Online]. Available: <https://books.google.com.br/books?id=SpvBd7IUCxkC>
- [29] I. Avciabas, B. Sankur, and K. Sayood, "Statistical evaluation of image quality measures," *Journal of Electronic Imaging*, vol. 15, no. 2, p. 206–223, Nov. 2002.
- [30] T. Landgrebe and R. Duin, "A simplified extension of the area under the roc to the multiclass domain," *Seventeenth Annual Symposium of the Pattern Recognition Association of South Africa*, 01 2006.