

UTILIZAÇÃO DE TÉCNICAS COMPUTACIONAIS PARA A DIFERENCIAÇÃO DE TECIDOS CÓLICOS EM IMAGENS COLONOSCÓPICAS

FERRERO CA; LEE HD; CHERMAN EA; COY CSR; GÓES JRN; WU FC; FAGUNDES JJ

Lab. de Bioinformática/Univ. Est. do Oeste do Paraná-UNIOESTE/PTI e Serviço de Coloproctologia-UNICAMP

OBJETIVOS: Propor uma metodologia apoiada por métodos de Análise de Imagens (AnI) e Inteligência Artificial (IA) para a utilização de atributos baseados em textura para caracterizar tecidos cólicos em imagens de exames de colonoscopia. **MÉTODOS:** neste trabalho é proposta uma metodologia para a AnI de exames de colonoscopia, constituída de quatro etapas: coleta do conjunto de imagens, construção do vetor de características, construção do modelo para classificação e avaliação do modelo. Etapa 1: consiste em selecionar, manualmente, para cada uma das imagens consideradas, duas classes de fragmentos: um de tecido cólico com alguma anormalidade, por exemplo pólipo, e outro de tecido cólico sem anormalidade. Etapa 2: são utilizados métodos de Análise de Imagens, para representa-las por um conjunto de características, as quais são comumente classificadas como baseadas em cor, forma e textura. Esse último permite obter informações sobre propriedades de disposição e espaço dos níveis de cinza que a compõem. O Espectro de Textura da imagem (ET), consiste na identificação de Unidades de Textura (UT) presentes na imagem. Para cada ponto r da imagem é construída uma matriz de ordem 3 na qual o elemento central é representado por r e os elementos adjacentes assumem valores 0, 1 ou 2 se o valor adjacente de r na imagem é menor, igual ou maior a r , respectivamente. Sendo possíveis 6561 UTs diferentes, o ET consiste em um histograma no qual o eixo das abscissas indica as UTs e o das ordenadas a freqüência de cada UT na imagem. Os histogramas são construídos comumente para cada componente da imagem (vermelho – R, verde – G, azul – B, intensidade, matiz e saturação). Etapa 3: após a extração de características, são aplicados métodos de IA, com o objetivo de construir padrões que permitam classificar novas imagens. Nesse sentido, um dos métodos de construção de padrões é o baseado em exemplos em que, dado um novo exemplo, são considerados os exemplos mais próximos (utilizando uma medida de distância) para tomar a decisão de atribuição da classe. Etapa 4: os modelos construídos são avaliados, primeiramente, através do método de Validação Cruzada (VC), que consiste na partição do conjunto de dados em k partições menores, em que cada partição é testada a partir do modelo construído com o restante das partições, obtendo uma aproximação do erro verdadeiro do modelo e, posteriormente, através da utilização da Tabela de Contingência (TC) para o cálculo de diversas métricas de precisão como Valor Preditivo Positivo (VPP), Valor Preditivo Negativo (VPN), Sensibilidade (S) e Especificidade (E). Foi construído um aplicativo computacional, denominado *Medical Image Analysis System*, para auxiliar em todas as etapas dessa metodologia. A metodologia apresentada foi aplicada a um conjunto de 67 imagens de exames de colonoscopia contendo pólipos, coletadas do Serviço de Coloproctologia da UNICAMP. Para cada imagem, foram selecionados dois fragmentos de tecido cólico, um contendo anormalidade do tipo pólipo e outro sem presença de anormalidade. Para os 134 fragmentos selecionados foram extraídos os ETs de R, G e B. Desse modo, cada fragmento foi representado por três conjuntos de 6561 elementos e um atributo que identifica a classe a que pertence – tecido de pólipo ou sem pólipo. Através do método de aprendizado baseado em exemplos, utilizando uma medida de similaridade baseada no teste estatístico Qui-Quadrado, foram determinados, para cada imagem, os ETs de cada cor mais próximos contidos no conjunto de treinamento. Os modelos construídos foram avaliados através do método de VC, utilizando dez partições, e através das medidas derivadas da TC. **RESULTADOS:** considerando o método de VC os erros médios e desvios-padrão para R, G e B foram respectivamente, 0,39 e 0,11, 0,29 e 0,08 e 0,28 e 0,11. O teste estatístico t para dados não emparelhados foi aplicado e obteve-se um p-valor $> 0,05$, indicando a não existência de diferença estatisticamente significativa entre os erros de classificação dos modelos construídos. Embora o erro médio obtido nesses experimentos esteja em torno de 30%, representou um ganho em relação ao erro da classe majoritária, a qual é de 50%. Para avaliar os modelos construídos foram ainda utilizadas as medidas relacionadas à tabela de contingência. O VPP foi de 65,96%, 78,00% e 79,59% para R, G e B, respectivamente, o VPN foi de 58,62%, 66,67% e 67,05% para cada componente, a S foi de 46,27%, 58,21% e 58,21% e, a E de 76,12%, 83,58% e 85,07%. Por meio da análise de VPP e S, pode ser constatado que, os modelos construídos têm dificuldade para identificar imagens de pólipos, porém quando uma imagem é classificada como sendo tecido de pólipo, a taxa de acerto é elevada. Por outro lado, quando analisados VPN e E, pode ser observado que, os modelos classificam uma grande parte das imagens como sendo tecidos sem pólipos, sacrificando assim, a taxa de acerto de classificação dessa classe. **CONCLUSÕES:** os resultados deste trabalho mostraram que, do ponto de vista de precisão, não houve diferença estatisticamente significativa entre a qualidade dos modelos construídos a partir dos ETs de R, G e B para os 134 fragmentos de tecido cólico considerados. Porém, uma análise através da TC, mostrou que os modelos construídos a partir das componentes G e B apresentaram maior probabilidade de predição, para as quatro medidas avaliadas, que o modelo construído a partir do ET de R. Esses modelos, tenderam a classificar imagens como sem pólipo, sendo a grande parte das imagens classificadas como de pólipos eram, de fato, tecidos cólicos de pólipos. Desse modo, foi observado que existem padrões pictóricos contidos nas imagens, no entanto outras características poderiam ser utilizadas e/ou propostas para cobrir adequadamente os exemplos que não foram cobertos utilizando o ET da imagem. Um dos trabalhos futuros inclui a coleta de um número maior de imagens tipo pólipo, a qual poderá auxiliar na construção de modelos mais robustos. Paralelamente, estão sendo implementados algoritmos para a extração de outras características que podem possibilitar a descoberta de outros atributos relevantes para descrever imagens de exames de colonoscopia.