

## **ESTUDO COMPARATIVO DE MODELOS COMPUTACIONAIS GERADOS SOBRE REPRESENTAÇÕES DE IMAGENS DE COLOSCOPIA: TECIDO DE MUCOSA NORMAL VS TECIDO DE MUCOSA DE PÓLIPO CÓLICO\***

FERRERO CA; LEE HD; SPOLAÔR N; COY CSR; FAGUNDES JJ; MACHADO RB; WU FC

Laboratório de Bioinformática/UNIOESTE/PTI e Serv. Coloproctologia/Faculdade de Ciências Médicas/UNICAMP

**OBJETIVOS:** Analisar a qualidade preditiva de modelos computacionais para a diferenciação de tecidos cólicos, construídos a partir da representação de Imagens de Coloscopia (IC) como Matrizes de Co-ocorrência (MC).

**MATERIAL E MÉTODOS:** Os modelos foram construídos aplicando técnicas de análise de imagens e de inteligência artificial em uma metodologia constituída de quatro etapas: coleta do conjunto de IC, construção do Vetor de Características (VC), construção de modelos e avaliação de modelos. O VC representa IC por meio de características de textura extraídas computacionalmente, as quais são calculadas utilizando MC, que descrevem as transições de intensidade em imagens. Cada posição da MC indica a frequência das transições considerando uma Distância (D), a qual permite identificar características em diferentes níveis de aproximação. Neste trabalho foram utilizadas 67 IC que apresentam pólipos, providas do Serviço de Coloproctologia da UNICAMP e coletadas no período de out/2004 a nov/2005. Foram extraídos uma parte de tecido de pólipo e outra de tecido sem pólipo adjacente, totalizando 134 partes. Para cada parte de tecido, foram criadas MC de D = 1, 2, 3, 4 e 5 e extraídas 11 características de textura, com as quais foram construídos cinco modelos computacionais utilizando o algoritmo J48 para a construção de árvores de decisão. Os modelos foram avaliados utilizando: (1) validação cruzada com dez partições e (2) tabelas de contingência, das quais foram extraídas as medidas Valor Preditivo Positivo (VPP), Valor Preditivo Negativo (VPN), Sensibilidade (S) e Especificidade (E).

**RESULTADOS:** Na análise (1), constatou-se que o modelo de D = 3 apresentou o menor erro médio, 22,25%, com desvio-padrão de 11,85%. A aplicação do teste *t* para dados emparelhados constatou diferença estatisticamente significativa, com nível de confiança de 95%, apenas entre os modelos de D = 3 e D = 4. A análise (2) indicou que o modelo de D = 1 apresentou os maiores valores de VPP e E, 80,7% e 83,58%, respectivamente. Os maiores valores de VPN e S foram, respectivamente, 78,46% e 79,1%, correspondentes ao modelo de D = 3.

**CONCLUSÃO:** As análises indicaram que os modelos D = 3 e D = 1 apresentaram as melhores qualidades preditivas. O modelo de D = 3 apresentou maiores precisões de VPN e S, e valores inferiores para VPP e E, em relação ao modelo de D = 1. Trabalhos futuros incluem a utilização de outras características para a representação de IC.

\*Trabalho apoiado pelo PDTA-FPTI/BR.