

# Utilização do Espectro de Textura para a Diferenciação de Imagens Cólicas em Exames Colonoscópicos

**Carlos Andres Ferrero<sup>1,3</sup>, Hwei Diana Lee<sup>1</sup>, Everton Alvares Cherman<sup>1</sup>,  
Cláudio Sady Rodrigues Coy<sup>2</sup>, João José Fagundes<sup>2</sup>,  
Juvenal Ricardo Navarro Góes<sup>2</sup>, Feng Chung Wu<sup>1,2</sup>**

<sup>1</sup>*Centro de Engenharias e Ciências Exatas – Universidade Estadual do Oeste do Paraná,  
Laboratório de Bioinformática – LABI, Parque Tecnológico Itaipu – PTI,  
Foz do Iguaçu, PR, Brasil*

<sup>2</sup>*Faculdade de Ciências Médicas – Universidade Estadual de Campinas,  
Serviço de Coloproctologia, Campinas, SP, Brasil*

<sup>3</sup>*Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo,  
Laboratório de Inteligência Computacional – LABIC, São Carlos, SP, Brasil*

## Resumo

*O armazenamento, cada vez mais freqüente, de exames através de imagens tem incentivado a aplicação de métodos de Análise de Imagens e de Inteligência Artificial para identificar padrões. Neste trabalho é apresentada uma metodologia para a caracterização de tecidos cólicos aplicada a um conjunto de 67 imagens de exames de colonoscopia. Essas imagens foram descritas pelos espectros de textura das componentes de cor: vermelho, verde e azul. Os modelos construídos permitiram evidenciar que a característica espectro de textura é adequada para, juntamente com outros atributos, realizar uma análise mais completa das imagens e, desse modo, obter padrões que auxiliem, mais eficientemente, especialistas nessa área.*

## Palavras Chave

Análise de Imagens, Espectro de Textura, Inteligência Artificial, Colonoscopia, Doenças do Intestino Grosso, Câncer Colorretal.

## 1. Introdução

Os sistemas de gerenciamento de dados têm permitido, nas diversas áreas de conhecimento, o acúmulo e a organização das informações bem como o crescimento de repositórios de dados multimídia (vídeo, imagens e texto) [1, 2, 3]. Decorrente ao fato, surgiu a necessidade de aplicar métodos para a análise desses dados e, conseqüentemente, prover suporte a processos de tomada de decisões. No contexto de dados do tipo imagem, a Análise de Imagens – AI – e a Inteligência Artificial – IA –, apresentam-se como recursos capazes de auxiliar nessa tarefa, a qual consiste na utilização de características visuais intrínsecas das imagens para a identificação de padrões que permitam classificar novos conjuntos de imagens.

Na área médica, imagens digitais são amplamente produzidas e utilizadas pelos especialistas para o registro e o diagnóstico de doenças. Uma das doenças de maior ocorrência no Brasil e no mundo é o câncer colorretal que, segundo o Instituto Nacional do Câncer, constitui a quarta maior incidência entre todos os tumores malignos, independentemente do sexo [4]. Em relação à mortalidade, as neoplasias do intestino grosso também ocupam o quarto lugar, ocorrendo em menor freqüência apenas que tumores de

pulmão, estômago e mama [5]. Para auxiliar no diagnóstico de doenças do intestino grosso, exames de colonoscopia apresentam-se como ferramentas indispensáveis [4, 5].

Desse modo, para auxiliar em processos de tomada de decisões associados às anormalidades evidenciadas por meio desse exame, podem ser aplicados métodos e ferramentas de Análise de Imagens e de Inteligência Artificial.

Nesse contexto, este trabalho constitui parte do desenvolvimento do projeto de Análise de Imagens Médicas, o qual está sendo desenvolvido em uma parceria entre o Laboratório de Bioinformática – LABI – da Universidade Estadual do Oeste do Paraná – UNIOESTE – e o Serviço de Coloproctologia da Faculdade de Ciências Médicas – FCM – da Universidade Estadual de Campinas – UNICAMP. Na Figura 1 estão representadas as fases constituintes do projeto de Análise de Imagens Médicas.

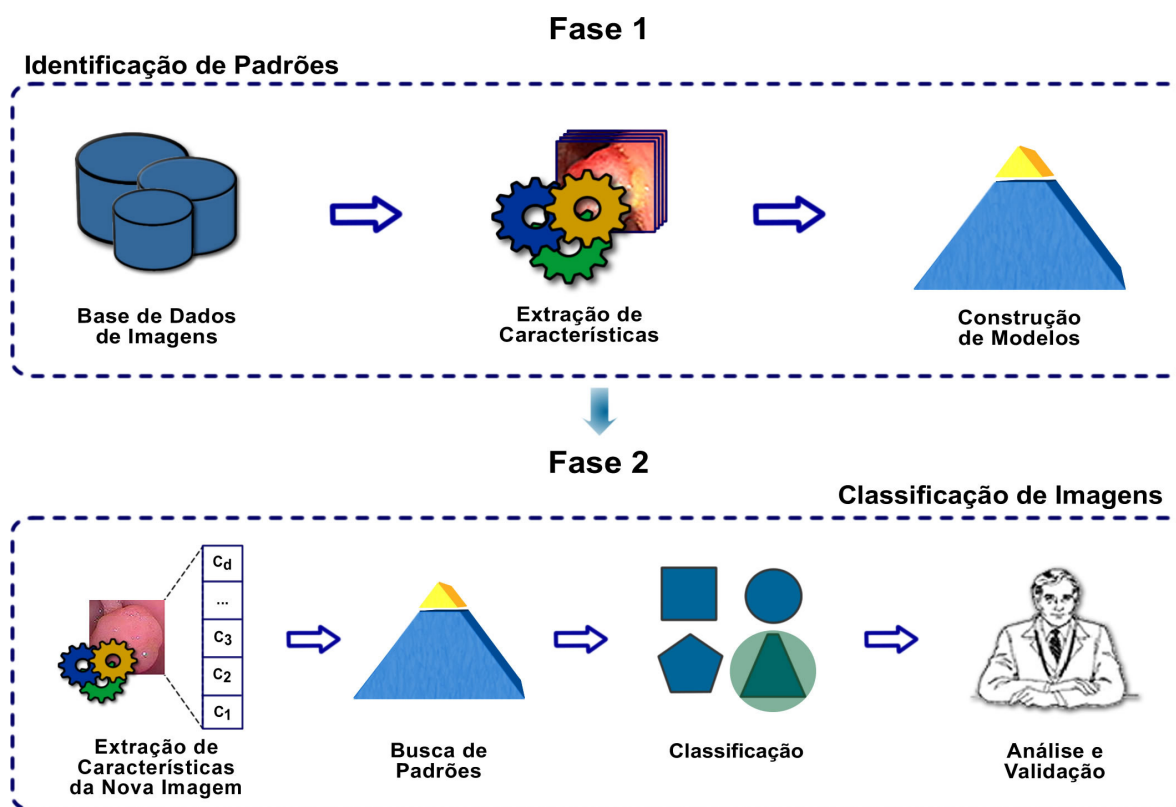


Figura 1: Fases que constituem o projeto de Análise de Imagens Médicas.

Na primeira fase, Identificação de Padrões, imagens contidas numa base de dados são representadas por um conjunto de características, a partir do qual são construídos modelos. Após, com base nesses modelos é executada a segunda fase, caracterizada como Classificação de Imagens, em que os modelos construídos são utilizados com a finalidade de desenvolver sistemas computacionais que permitam classificar novas imagens e, desse modo, auxiliar especialistas nos processos de tomada de decisões.

Neste trabalho é apresentada uma metodologia para a utilização de atributos baseados em textura para a caracterização de tecidos cólicos e está estruturado da seguinte

maneira. Na Seção 2 é descrita a metodologia proposta e na Seção 3 é apresentado o aplicativo construído para implementar essa metodologia. Na Seção 4 é mostrado um estudo de caso utilizando um conjunto de imagens de exames de colonoscopia e na Seção 5 são apresentadas as conclusões e os trabalhos futuros.

## 2. Metodologia Proposta

A metodologia proposta tem como objetivo a utilização de atributos baseados em textura para a caracterização de tecidos cólicos em imagens de exames de colonoscopia e está constituída de quatro etapas: coleta do conjunto de imagens, construção do vetor de características, construção do modelo para classificação e avaliação do modelo. Essas etapas são descritas a seguir.

### 2.1. Etapa 1: Coleta do Conjunto de Imagens

Hospitais e clínicas médicas utilizam sistemas computacionais para o registro de exames de colonoscopia através de imagens e laudos médicos. Essas imagens contêm informações sobre diversos tipos de anormalidade como polipóide, diverticular e vascular entre outras. Com isso, é importante a diferenciação entre tecidos correspondentes a anormalidades e tecidos sem anormalidade. Essa etapa consiste em selecionar manualmente, para cada uma das imagens consideradas, um fragmento de tecido cólico com alguma anormalidade e um fragmento de tecido cólico que não apresente anormalidade. Um exemplo de seleção de tecidos cólicos é ilustrado na Figura 2, a qual apresenta uma imagem real de exame de colonoscopia contendo uma anormalidade do Tipo Ip (Pólipo Protruso Pediculado, segundo a classificação morfológica de pólipos da Sociedade Japonesa de Pesquisa do Câncer Colorretal) [4, 5]. Nessa imagem, são ressaltadas as duas classes de fragmentos, *a* e *b*, que representam tecidos cólicos com e sem presença de anormalidade, respectivamente.

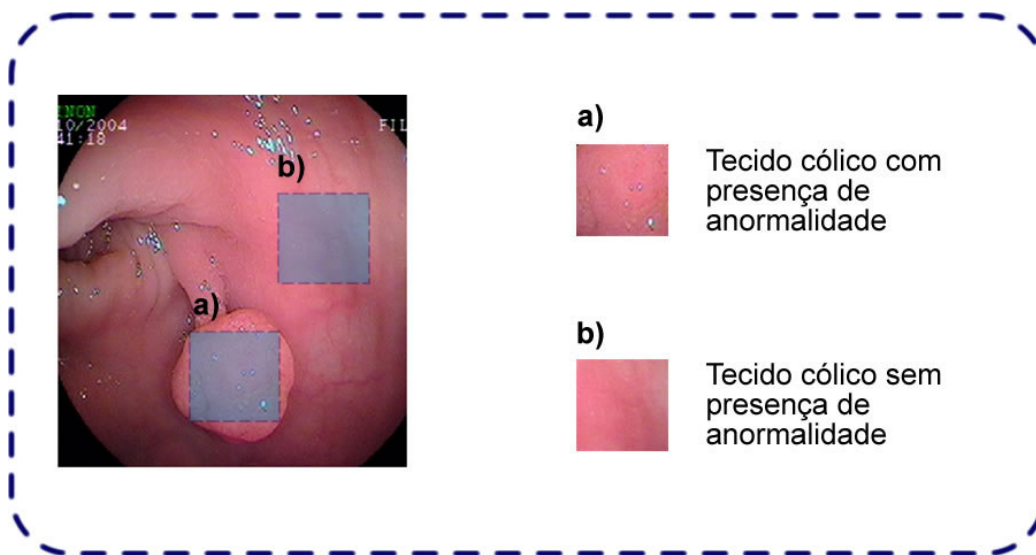


Figura 2: Exemplo de coleta de imagens através da seleção de um tecido de pólipo e um tecido sem anormalidade.

Uma vez realizada a coleta dos dois fragmentos para cada uma das imagens consideradas é constituído o Conjunto de Imagens – CI – definido por

$CI = \{Im_1, Im_2, \dots, Im_n, Im_{n+1}, Im_{n+2}, \dots, Im_{2n}\}$ , em que cada elemento de  $CI$  representa um registro de imagem e  $n$  o número de imagens consideradas.

## 2.2. Etapa 2: Construção do Vetor de Características

No contexto de Análise de Imagens, imagens são representadas por meio de um conjunto de características, as quais são comumente classificadas como baseadas em cor, textura e forma. Essa extração de características permite que as imagens sejam entendidas não apenas como um conjunto de pontos (pixels) mas como um grupo de informações em que cada uma delas tem um valor associado. Desse modo, é possível obter uma representação da imagem no formato atributo-valor, que é o formato de entrada de dados comumente utilizado pela maioria dos algoritmos para a construção de padrões.

Nesse sentido, após a coleta de imagens é realizada a extração de características. Cada imagem é representada por um vetor de características  $V = \{C_1, C_2, \dots, C_d\}$ , onde  $d$  corresponde à dimensão do vetor, isto é, ao número de características extraídas.

Cada exemplo (imagem) é descrito pelo conjunto  $E_i = \{A_{i,1}, A_{i,2}, \dots, A_{i,d+1}\}$  composto por  $d+1$  valores de atributos, no qual o último atributo indica a classe a que o exemplo pertence e  $A_{i,j}$  corresponde ao valor do atributo (característica)  $j$  do exemplo  $i$ . Com base em  $CI$  é construído o Conjunto de Dados Inicial – CDI – definido como  $CDI = \{E_1, E_2, \dots, E_{2n}\}$ , isto é,  $2n$  exemplos descritos por  $d+1$  atributos.

### 2.2.1. Características Baseadas em Textura

Atributos baseados em textura constituem um dos aspectos mais importantes da percepção visual humana. Essas características, permitem obter informações sobre propriedades de disposição e espaço dos principais elementos da imagem, por exemplo, os níveis de cinza que a compõem. Uma das formas de extrair esses atributos é através do Espectro de Textura da imagem – ET. Nesse contexto, um dos métodos mais utilizados para extração de características, baseado no ET da imagem, consiste na identificação de Unidades de Textura – UT [6]. Para cada pixel  $r$  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  $r$  é menor, igual ou maior a  $r$ , respectivamente. Na Figura 3 é ilustrada a construção de uma UT a partir dos pixels adjacentes a um pixel  $r$ .

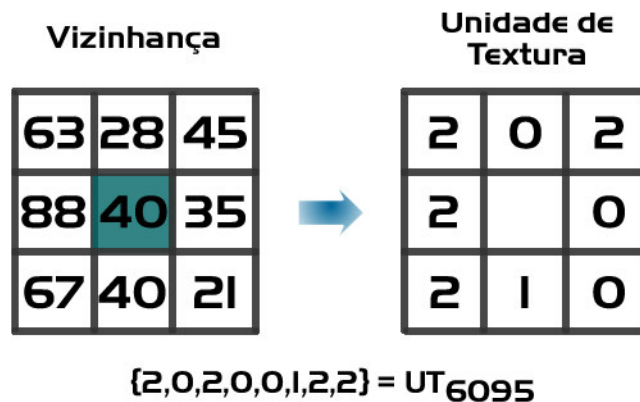


Figura 3: Representação da vizinhança correspondente ao pixel  $r$  em uma unidade de textura.

Desse modo, são possíveis 6561 UTs diferentes. Em razão disso, o ET consiste em um histograma no qual o eixo das abscissas indica as UTs e o das ordenadas a frequência de cada UT na imagem. Os histogramas são construídos comumente a partir da imagem no formato de níveis de cinza [1, 6] ou para cada componente de cor da imagem: R, G, B, I, H, S<sup>1</sup> [7]. Com isso, cada imagem pode ser descrita pelo espectro de textura de cada componente, sendo representada por  $E_i = \{A_{i,1}, A_{i,2}, \dots, A_{i,6561}, A_{i,6562}\}$ , em que o atributo  $A_{i,UT}$  indica a frequência de cada unidade de textura dentro da imagem  $i$  e o último atributo refere-se à classe a que o exemplo pertence.

### 2.3. Etapa 3: Construção de Modelos para Classificação

Após a extração de características, é executada a etapa de construção de modelos para classificação de exemplos. Para isso, podem ser aplicados métodos de IA, com o objetivo de construir sistemas capazes de adquirir conhecimento. Esses sistemas consistem em programas de computador que podem tomar decisões com base na experiência acumulada através da solução bem-sucedida de problemas anteriores [8]. Desse modo, podem ser utilizados algoritmos de IA para construir modelos a partir de grandes conjuntos de dados com o intuito de encontrar padrões que permitam classificar exemplos futuros. Nesse contexto, a partir de *CDI*, que contém  $2n$  exemplos descritos pelo espectro de textura de cada componente e pela sua classe, é realizada a construção de modelos a partir desses dados.

#### 2.3.1. Paradigma Baseado em Exemplos

Uma maneira de realizar classificação de novas situações é através da similaridade com outras experiências. Nesse sentido, os sistemas que utilizam esses métodos não paramétricos de aprendizado são denominados de baseados em exemplos, ou memória, e consistem na comparação de cada novo caso com exemplos de treinamento previamente armazenados em memória. Considerando-se o(s) exemplo(s) mais próximo(s) é realizada a tomada de decisão de atribuição da classe ao novo exemplo [9,10]. Nesse tipo de aprendizado, o conjunto de exemplos de treinamento constitui o modelo de classificação.

Dentro desse paradigma, têm-se três características importantes que devem ser consideradas: quais casos conhecidos devem ser lembrados, qual deve ser a medida que quantifica quão similar um novo exemplo é dos contidos na base de dados e de que modo um novo caso será relacionado com os casos armazenados. Essas características, as quais influenciam fortemente a *performance* do classificador, são apresentadas a seguir:

- *Tamanho do conjunto de treinamento*: se todos os exemplos forem memorizados, a classificação pode se tornar demorada e difícil de manter. Em vez disso, o ideal é armazenar apenas os casos mais representativos de cada classe, resumindo a informação mais importante em um conjunto menor de exemplos. Esse número de exemplos que irá constituir o modelo, define também o custo em tempo para localizar o exemplo mais similar, a complexidade do algoritmo utilizado e o espaço em memória para armazenar os casos [9,10].
- *Medida de similaridade*: uma das questões mais relevantes relacionada ao processo de classificação é a escolha da medida utilizada para determinar o grau de

---

<sup>1</sup> Red, Green, Blue, Intensity, Hue, Saturation.

similaridade entre um exemplo e cada um dos exemplos contidos no conjunto de treinamento, para que esse novo exemplo possa ser rotulado [9]. Em um problema de atributos com valores numéricos, uma maneira de medir a distância entre dois exemplos é por meio do cálculo da raiz quadrada da soma dos quadrados da diferença entre atributos de cada exemplo, a qual é denominada de distância Euclidiana [12,13]. Essa e outras medidas, como Manhattan, Mahalanobis e Qui-Quadrado são utilizadas, respectivamente, para quantificar a distância absoluta entre os exemplos, para medir a similaridade quando os valores dos atributos encontram-se em escalas diferentes e para calcular o grau de similaridade quando os valores dos atributos são dados numéricos que não correspondam a uma distribuição normal. Nesse aspecto, uma questão importante é a existência de atributos não importantes, que poderia resultar na conclusão de que dois exemplos que apresentam muita diferença em um subconjunto de atributos não importantes sejam diferentes o suficiente para não considerá-los no momento da tomada de decisão, sendo que poderiam pertencer à mesma classe. Desse modo, existe um conjunto de atributos que influenciam de maneira desfavorável a *performance* do classificador.

- *Cardinalidade do relacionamento entre os exemplos*: a determinação da classe que será atribuída a um exemplo não rotulado pode ser realizada de duas maneiras considerando-se os vizinhos mais próximos. A primeira é baseada na classe de um único exemplo mais próximo encontrado, considerado uma medida de similaridade, no conjunto de treinamento. A segunda maneira de rotular um novo exemplo é baseada na análise de diversos vizinhos mais próximos, tomando a decisão em função da classe majoritária entre esses exemplos [9].

#### 2.4. Etapa 4: Avaliação de Modelos

Na quarta etapa, os modelos construídos são avaliados através de métodos para a avaliação de modelos, com o intuito de analisar a qualidade dos padrões encontrados. Entre as diversas maneiras de se avaliar modelos têm-se o de validação cruzada e a utilização da Tabela de Contingência – TC – para o cálculo de diversas métricas de precisão como valor preditivo positivo, valor preditivo negativo, sensibilidade e especificidade. Ainda nessa etapa, os modelos construídos são avaliados e validados em conjunto com especialistas do domínio.

Essas avaliações podem contribuir não apenas para quantificar a qualidade das características extraídas e dos modelos construídos, mas também para a escolha e a investigação de novas características presentes nas imagens que poderiam ser extraídas computacionalmente.

##### 2.4.1. Validação Cruzada

Um dos métodos que pode auxiliar na avaliação e validação de modelos é o de validação cruzada (*Cross-Validation*). Esse método consiste em dividir aleatoriamente um conjunto de exemplos em  $k$  amostras iguais, usualmente dez, e criar  $k$  grupos, onde a  $k$ -ésima amostra é o conjunto de exemplos de teste e as  $k-1$  amostras restantes constituem o conjunto de exemplos de treinamento. Para cada grupo é construído um modelo com os exemplos de treinamento e testado com o conjunto de exemplos de teste. Desse modo, o erro médio do classificador é estimado como a média dos erros dos  $k$  modelos construídos [10].

#### 2.4.2. Tabela de Contingência

Em estatística, as tabelas de contingência são utilizadas para avaliar o relacionamento entre duas ou mais variáveis nominais, isto é, para identificar se pertencem ou não a uma categoria. Por exemplo, neste trabalho trata-se da classificação ou não de um fragmento de tecido cólico como sendo um tecido de pólipos ou sem pólipos. Na Tabela 1 é apresentado um exemplo de tabela de contingência. Nessa tabela, as variáveis *A*, *B*, *C* e *D*, indicam o número de verdadeiros positivos, falsos positivos, falsos negativos e verdadeiros negativos, respectivamente.

Tabela 1: Tabela de contingência.

Tecido	Classificação		Totais
	De Pólipo	Sem Pólipo	
De Pólipo	A	C	A+C
Sem Pólipo	B	D	B+D
Totais	A+B	C+D	A+C+B+D

Considerando a classe positiva como as imagens de fragmentos de tecido cólico de pólipos e a classe negativa como as imagens de tecido cólico sem pólipos, a partir da tabela de contingência podem ser extraídas as seguintes medidas de avaliação:

- *Valor Preditivo Positivo* – VPP: percentagem de exemplos da classe positiva do total de exemplos classificados como positivos. Calculado por:  $VPP = A/(A + B)$ .
- *Valor Preditivo Negativo* – VPN: percentagem de exemplos da classe negativa do total de exemplos classificados como negativos. Calculado por:  $VPN = D/(C + D)$ .
- *Sensibilidade*: percentagem de exemplos classificados como positivos do total de exemplos positivos. Calculado por:  $Sensibilidade = A/(A + C)$ .
- *Especificidade*: percentagem de exemplos classificados como negativos do total de exemplos negativos. Calculado por:  $Especificidade = D/(B + D)$ .

### 3. Aplicativo *Medical Image Analysis System*

Para auxiliar na aplicação da metodologia anteriormente apresentada foi construído um sistema computacional denominado *Medical Image Analysis System* – MIAS – [14], que foi desenvolvido em linguagem de programação JAVA<sup>2</sup> no ambiente de desenvolvimento Eclipse<sup>3</sup> e utilizando a biblioteca *Java Advanced Imaging* – JAI – [15]. O aplicativo provê suporte tanto aos processos de seleção, operação, manipulação e visualização de imagens, como também à extração de características intrínsecas da imagem. O sistema MIAS – Figura 4 – está constituído pelos módulos de visualização de diretórios, de listagem de imagens, de visualização de imagens e de extração de características.

<sup>2</sup> <http://www.java.sun.com/>

<sup>3</sup> <http://www.eclipse.org/>.

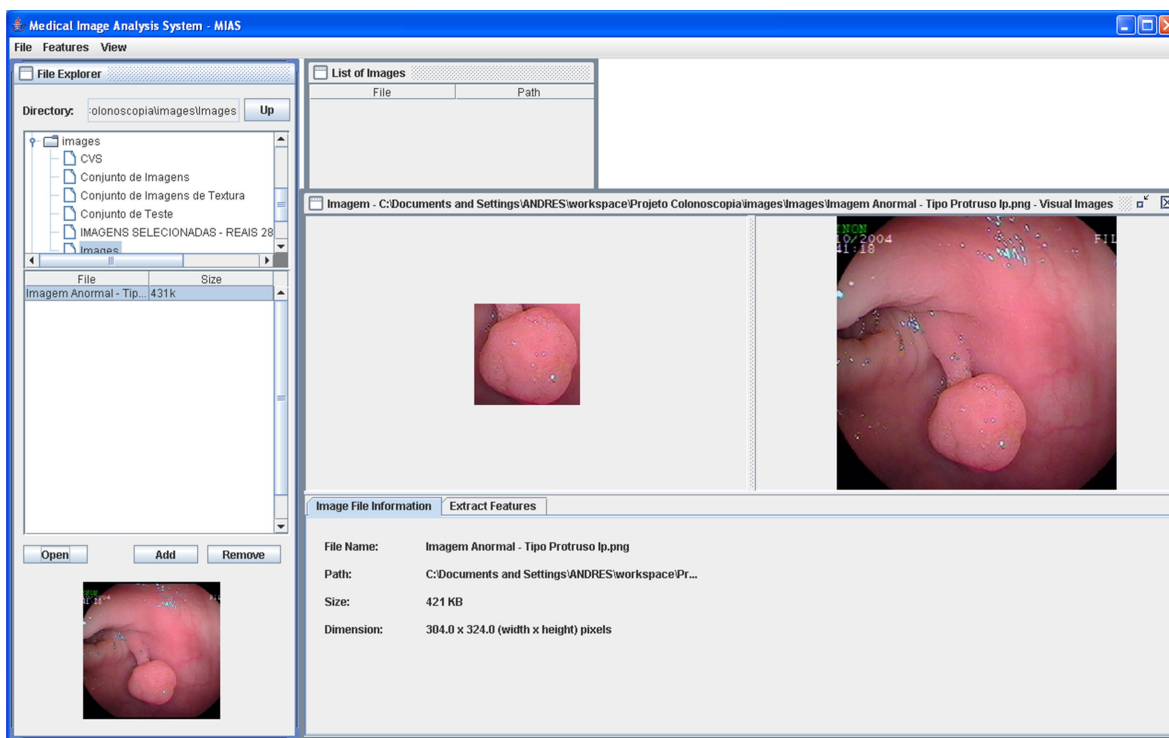


Figura 4: Screenshot do aplicativo MIAS.

#### 4. Aplicação da Metodologia Proposta em Imagens de Exames de Colonoscopia

A metodologia apresentada na Seção 2 foi aplicada a um conjunto de 67 imagens de exames de colonoscopia contendo pólipos, coletadas do Serviço de Coloproctologia da FCM/UNICAMP. Inicialmente, para cada imagem, foi realizada a seleção dos fragmentos de tecido com e sem presença de pólipo. Um exemplo dessa seleção de tecidos cólicos é apresentado na Figura 2. Para a realização dessa tarefa foi utilizado o aplicativo MIAS e o conjunto de imagens  $CI = \{Im_1, Im_2, \dots, Im_{134}\}$  foi construído.

Na segunda etapa, para a representação das imagens foi utilizado o método para a extração do espectro de textura, que foi apresentado na Seção 2.2.1 e consiste em quantificar a frequência de unidades de textura, considerando as diferentes componentes da imagem [6,7]. Neste trabalho foi realizada a extração do ET para as componentes  $R$ ,  $G$  e  $B$ , permitindo descrever cada imagem de  $CI$  por três conjuntos do tipo  $V = \{UT_1, UT_2, \dots, UT_{6561}\}$ , correspondentes a cada componente [16]. Para a representação das imagens foi utilizado o módulo de extração de características do aplicativo MIAS, no qual as imagens foram convertidas para 64 níveis de cinza, a qual é uma representação utilizada comumente na extração de características baseadas em textura a partir de imagens médicas [2,17]. Desse modo, foi constituído o conjunto de dados, definido por  $CD = \{E_1, E_2, \dots, E_{134}\}$ , em que cada elemento  $E_i$  é representado por três conjuntos de 6561 elementos e um atributo que identifica a classe a que pertence.

Posteriormente, foi realizada a etapa de construção de modelos. Nessa etapa, foi utilizado o paradigma de aprendizado baseado em exemplos, apresentado na Seção 2.3.1, em que o conjunto de exemplos de treinamento representa o próprio modelo [9]. Com base nesse paradigma, o conjunto de exemplos de teste é classificado utilizando uma medida de



similaridade baseada no teste estatístico Qui-Quadrado, para determinar o espectro de textura mais próximo contido no conjunto de treinamento.

Para avaliar o modelo construído foi utilizada primeiramente a técnica de validação cruzada, mencionada na Seção 2.4.1, em que o conjunto de dados CD foi dividido em dez partições. Na Tabela 2 são apresentados o erro médio e o desvio-padrão de classificação para cada componente analisada.

*Tabela 2: Validação Cruzada para avaliar o modelo construído.*

	<b>R</b>	<b>G</b>	<b>B</b>
<b>Erro Médio</b>	0,3879	0,2912	0,2830
<b>Desvio Padrão</b>	0,1125	0,0824	0,1133

De acordo com essas estimativas, pode ser observado que o espectro de textura da componente B teve o menor erro médio. É importante ressaltar que, apesar do espectro de textura G não ter apresentado o menor erro médio, apresentou o menor desvio padrão, demonstrando que sofreu a menor variação no erro médio. Com base nos erros de classificação de cada componente para cada uma das partições, os quais passaram no teste da normalidade, foi aplicado um teste estatístico paramétrico para dados não emparelhados, com o intuito de constatar a existência de diferença estatisticamente significativa entre os erros de classificação. O *p-valor* obtido foi maior que *0,05*, indicando que não existe diferença estatisticamente significativa entre os erros de classificação dos modelos de R, G e B. É interessante ressaltar que, embora o erro médio obtido nesses experimentos seja próximo a 30%, mesmo assim 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. Essas medidas são: valor preditivo positivo, valor preditivo negativo, sensibilidade e especificidade. Nas Tabelas 3, 4 e 5, são apresentadas as tabelas de contingência dos modelos construídos a partir da análise de R, G e B, respectivamente e, na Tabela 6 são mostrados os valores das medidas anteriormente citadas.

*Tabela 3: Tabela de contingência do modelo construído a partir da componente R.*

<b>Tecido</b>	<b>Classificação</b>		<b>Totais</b>
	<b>De Pólipo</b>	<b>Sem Pólipo</b>	
<b>De Pólipo</b>	31	36	67
<b>Sem Pólipo</b>	16	51	67
<b>Totais</b>	47	87	134

*Tabela 4: Tabela de contingência do modelo construído a partir da componente G.*

<b>Tecido</b>	<b>Classificação</b>		<b>Totais</b>
	<b>De Pólipo</b>	<b>Sem Pólipo</b>	
<b>De Pólipo</b>	39	28	67
<b>Sem Pólipo</b>	11	56	67
<b>Totais</b>	50	84	134

Tabela 5: Tabela de contingência do modelo construído a partir da componente B.

Tecido	Classificação		Totais
	De Pólipo	Sem Pólipo	
De Pólipo	39	28	67
Sem Pólipo	10	57	67
Totais	49	85	134

Tabela 6: Medidas calculadas a partir da TC dos modelos construídos.

Modelo	VPP	VPN	Sensibilidade	Especificidade
<b>R</b>	65,96	58,62	46,27	76,12
<b>G</b>	78,00	66,67	58,21	83,58
<b>B</b>	79,59	67,05	58,21	85,07

Os dados apresentados indicam que os modelos construídos a partir dos espectros de textura de cada componente apresentaram as seguintes características:

- A taxa de predição tecido de pólipo (VPP) foi maior para as componentes G e B;
- A taxa de predição de tecido sem pólipo (VPN) foi considerada baixa, para as três componentes, sendo que os modelos com maior VPN foram os correspondentes às componentes G e B;
- A probabilidade de que uma imagem, de tecido cólico de pólipo seja classificada, efetivamente como de pólipo (sensibilidade) foi considerada baixa para as três componentes, sendo os modelos mais precisos os das componentes G e B;
- A probabilidade de que uma imagem, de tecido cólico sem pólipo, seja efetivamente classificada como sem pólipo (especificidade) foi maior também para as componentes G e B, sendo que essa última apresentou o maior valor.

Com base nessas informações, pode ser observado que as componentes G e B apresentaram espectros de textura mais representativos para as classes de imagens consideradas – tecidos cólicos de pólipo e sem pólipo.

Por meio da análise do VPP e a sensibilidade, pode ser constatado que, do total de imagens de tecido de pólipo, apenas 58,21% foram classificadas como tais e mais de 78% (considerando os modelos de G e B) das imagens que os modelos classificaram como de pólipos, eram efetivamente tecidos de pólipo. Isto indica que o modelo construído tem 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 o VPN e a especificidade, pode ser observado que, considerando as imagens de tecido sem pólipo, mais de 83% foram classificadas corretamente para os modelos das componentes G e B, porém do total de imagens de fragmentos de tecidos sem pólipo classificadas, cerca de 66% foram classificadas corretamente. Isso comprova que o modelo construído classifica uma grande parte das imagens como sendo tecidos sem pólipos, sacrificando assim, a taxa de acerto de classificação dessa classe.

## 5. Conclusão

Os resultados deste trabalho mostraram que, do ponto de vista de precisão, não foi possível constatar uma diferença estatisticamente significativa entre a qualidade dos modelos construídos a partir dos espectros de textura das componentes R, G e B dos 134 fragmentos de tecido cólico considerados. Porém, uma análise através da tabela de contingência, mostrou que os modelos construídos a partir de G e B apresentaram maior probabilidade de predição, para as quatro medidas avaliadas, que o modelo construído a partir do espectro de textura de R. Esses modelos, tenderam a classificar imagens como sem pólipos, 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 espectro de textura da imagem.

Um dos trabalhos futuros inclui a coleta de um número maior de imagens de tipos bem definidos de anormalidades, a qual poderá auxiliar na construção de modelos mais adaptados para conjuntos maiores de imagens, além de permitir que padrões relevantes sejam encontrados. 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. Outro trabalho futuro inclui a aplicação de outros métodos de construção de padrões que permitam encontrar modelos mais relevantes e de maior precisão.

### Agradecimentos

Ao Programa de Desenvolvimento Tecnológico Avançado – PDTA/FPTI-BR – pelo auxílio por meio da linha de financiamento de bolsas.

### Referências

- [1] S. Karkanis, K. Galoussi e D. Maroulis, “Classification of endoscopic images based on texture spectrum”, *Advance Course in Artificial Intelligence (ACAI99)*, In *Proceedings of Workshop on Machine Learning in Medical Applications*, p. 63–69, 1999.
- [2] J. C. Felipe, A. J. M. Traina e C. Traina, “Retrieval by content of medical images using texture for tissue identification”, In *Proceedings of 16th IEEE Symposium on Computer-based Medical Systems (CBMS 2003)*, p. 26–27, 2003.
- [3] C. A. Ferrero, A. G. Maletzke, H. D. Lee, F. C. Wu, A. P. Neto, J. J. Fagundes e J. R. N. Góes, “Endia - uma ferramenta para a construção de repositórios de dados de endoscopia digestiva alta”, *X Congresso Brasileiro de Informática em Saúde (CBIS 2006)*, p. 1–6, 2006.
- [4] F. A. Quilici, *Colonoscopia. Lemos-Editorial*, São Paulo – SP, Brasil, 2000.
- [5] R. S. Cotran, V. Kumar e T. Collins, *Patologia Estrutural e Funcional*, Guanabara Koogan, 6 ed., Rio de Janeiro – RJ, Brasil, 2000.
- [6] D. C. He e L. Wang, “Texture unit, texture spectrum, and texture analysis”, In *Proceedings of IEEE Trans. on Geoscience and Remote Sensing*, vol. 28, n. 4, p. 509–512, 1990.
- [7] M. P. Tjoa e S. M. Krishnan, “Feature extraction for the analysis of colon status from the endoscopic images”, *BioMedical Engineering OnLine*, vol. 2, n. 9, p. 1-17 2003.
- [8] I. H. Witten e E. Frank, *Data mining: practical machine learning tools and techniques*, Elsevier, 2 ed., 2005.

- [9] M. C. Monard, G. E. A. P. A. Batista, S. Kawamoto e J. B. Pugliesi, “Uma introdução ao aprendizado de máquina por exemplos”, *Notas de aula do Instituto de Ciências Matemáticas de São Carlos – ICMC/USP*, São Carlos – SP, Brasil, 1997.
- [10] E. A. Elman, *Introduction to Machine Learning*, MIT Press, Cambridge – MA, England, 2004.
- [11] D. W. Aha, D. Kibler e M. K. Albert, “Instance-based learning algorithms”, *Machine Learning*, vol. 6, p. 37–66, 1991.
- [12] M. Kokare, B.N. Chatterji e P.K. Biswas, “Comparison of similarity metrics for texture image retrieval”, *Conference on Convergent Technologies for Asia-Pacific Region*, vol. 2, p. 571–575, 2003.
- [13] C. C. Chen, H. T. Chu. “Similarity measurement between images”, *Computer Software and Applications Conference (COMPSAC), 29th Annual International*, vol. 1, p. 41–42, 2005.
- [14] C. A. Ferrero, H. D. Lee, F. C. Wu, C. S. R. Coy, J. J. Fagundes e J. R. N. Góes, “Identificação de padrões de anormalidades em tecidos cólicos”, *I Congresso da Academia Trinacional de Ciências*, p. 1–1, 2006.
- [15] L. H. Rodrigues, *Building imaging applications with java technology*, Addison Wesley, Boston – MA, USA, 2001.
- [16] C. A. Ferrero, H. D. Lee, E. A. Cherman, C. S. R. Coy, J. J. Fagundes, J. R. N. Góes e F. C. Wu, “Utilização de Atributos Baseados em Textura para a Caracterização de Tecidos Cólicos em Imagens de Colonoscopia”, *VII Workshop de Informática Médica*, a ser publicado, 2007.
- [17] P. Howarth, A. Yavlinsky, D. Heesch e S. Rüger, “Medical image retrieval using texture, locality and colour”, *Lecture Notes from the Cross Language Evaluation Forum*, p. 740–749, 2005.
- [18] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann, San Francisco – CA, USA, 1993.
- [19] H. Motulsky, *Intuitive Biostatistics*, Oxford University Press, New York – NY, USA, 1995.

**Dados de Contato:**

Carlos Andres Ferrero<sup>1</sup> – [anfer@icmc.usp.br](mailto:anfer@icmc.usp.br)  
 Huei Diana Lee<sup>2</sup> – [huei@unioeste.br](mailto:huei@unioeste.br)  
 Everton Alvares Cherman<sup>3</sup> – [evertoncherman@gmail.com](mailto:evertoncherman@gmail.com)  
 Cláudio Sady Rodrigues Coy<sup>4</sup> – [ccooy@terra.com.br](mailto:ccooy@terra.com.br)  
 João José Fagundes<sup>5</sup> – [jffagundes@mpcnet.com](mailto:jffagundes@mpcnet.com)  
 Juvenal Ricardo Navarro Góes<sup>6</sup> – [rgoes@mpcnet.com.br](mailto:rgoes@mpcnet.com.br)  
 Feng Chung Wu<sup>7</sup> – [wufc@unioeste.br](mailto:wufc@unioeste.br)

<sup>1,2,3,7</sup> Centro de Engenharias e Ciências Exatas – Universidade Estadual do Oeste do Paraná, Laboratório de Bioinformática – LABI, Parque Tecnológico Itaipu – PTI, Caixa Postal 39, 85856-970 – Foz do Iguaçu, PR, Brasil.

<sup>4,5,6,7</sup> Faculdade de Ciências Médicas – Universidade Estadual de Campinas, Serviço de Coloproctologia, Caixa Postal 6111, 13083-970 – Campinas, SP, Brasil.

<sup>1</sup> Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo, Laboratório de Inteligência Computacional – LABIC, Caixa Postal 668, 13560-970 – São Carlos, SP, Brasil.