

## Utilização de Atributos Baseados em Textura para a Caracterização de Tecidos Cólicos em Imagens de Colonoscopia \*

Carlos Andres Ferrero<sup>1,3</sup>, Huei 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>, Wu Feng Chung<sup>1</sup>

<sup>1</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>2</sup>Faculdade de Ciências Médicas – Universidade Estadual de Campinas  
Serviço de Coloproctologia  
Caixa Postal 6111, 13083-970 – Campinas, SP, Brasil

<sup>3</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

{anfer86,hueidianalee,evertoncherman}@gmail.com

**Abstract.** *Application of Image Analysis and Artificial Intelligence methods to identify patterns has been stimulated by the increasing of storage of exams using images. In this work, it is presented a methodology to assist the characterization of colic tissues which was applied to a set of 15 images of colonoscopy exams. These images were described by 11 texture attributes from which two were found important according to the considered importance criterion. The constructed model showed that the selected attributes are good candidates to be used, in conjunction with other attributes, to perform a more complete image analysis and thus, to obtain patterns to give more efficient support to specialists in this domain.*

**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 15 imagens de exames de colonoscopia. Essas imagens foram descritas por 11 atributos de textura, dos quais dois apresentaram-se importantes segundo o critério considerado. Os modelos construídos permitiram evidenciar que as características selecionadas são boas candidatas para, juntamente com outros atributos, realizar uma análise mais completa das imagens e, desse modo, obter padrões que auxiliem de modo mais eficiente especialistas nessa área.*

---

\*Trabalho desenvolvido com o apoio da Fundação Parque Tecnológico Itaipu – FPTI – e do Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq.

## 1. Introdução

Os sistemas de gerenciamento de dados têm permitido, nas diversas áreas de conhecimento, o acúmulo e a organização das informações e o crescimento de repositórios de dados multimídia (vídeo, imagens e texto) [1, 2]. Decorrente desse fato, surgiu a necessidade de aplicar métodos para a análise desses dados e, desse modo, prover suporte a processos de tomada de decisões nas diversas áreas de conhecimento. 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 é o câncer colorretal que, segundo o Instituto Nacional do Câncer, constitui a quarta maior incidência entre todos os tumores malignos, independentemente do sexo [3]. Em relação à mortalidade, as neoplasias do intestino grosso também ocupam o quarto lugar, ocorrendo em menor frequência apenas que tumores de pulmão, estômago e mama [4]. Exames de colonoscopia são indispensáveis para o diagnóstico de doenças do intestino grosso [3, 4].

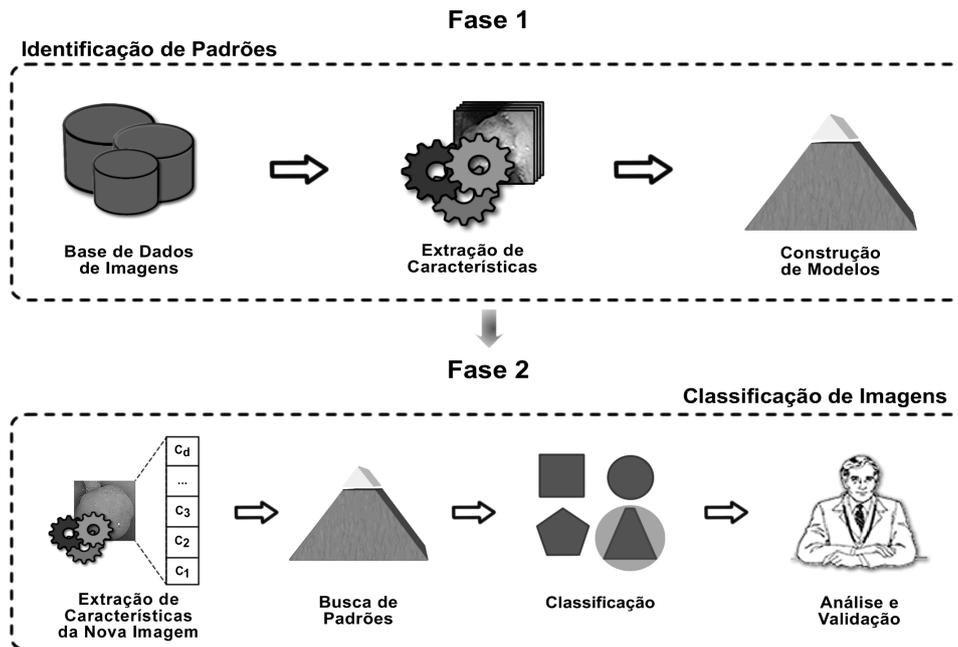
Desse modo, para auxiliar em processos de tomada de decisões associados a anormalidades evidenciadas por meio desse exame, podem ser aplicados métodos e ferramentas de Análise de Imagens e de Inteligência Artificial. Nesse contexto, este trabalho constitui parte do desenvolvimento do projeto de Análise 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.

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 cinco etapas: coleta do conjunto de imagens, construção do vetor de carac-



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

terísticas, seleção de características importantes, construção do modelo para classificação e avaliação do modelo. Essas etapas são descritas a seguir.

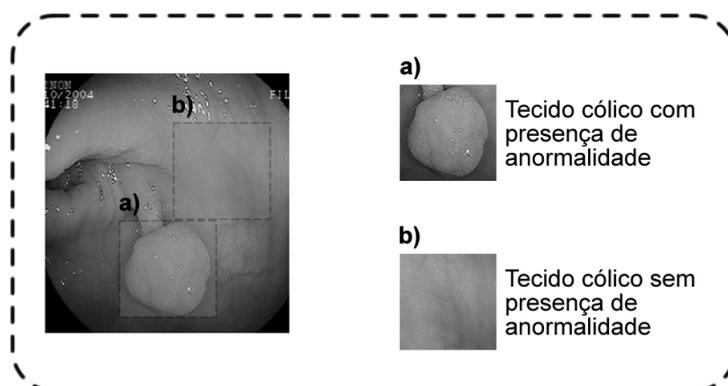
### 2.1. Etapa 1: Coleta do Conjunto de Imagens

Hospitais e clínicas médicas utilizam sistemas computacionais para o registro de exames de colonoscopia através de imagens. 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 (Protruso Pediculado, segundo a classificação morfológica de pólipos da Sociedade Japonesa de Pesquisa do Câncer Colorretal) [3, 4]. Nessa imagem, são ressaltadas as duas classes de fragmentos, *a* e *b*, que representam tecidos cólicos com e sem presença de pólipos, respectivamente.

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



**Figura 2. Exemplo de coleta de imagens através da seleção de um tecido de pólipó e um tecido com ausência de pólipó.**

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  $V = \{C_1, C_2, C_3, \dots, C_d\}$ , 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, sendo que, comumente esses atributos consistem na análise de um conjunto de matrizes de co-ocorrência [5, 6]. Essas matrizes representam as transições de níveis de cinza presentes na imagem, em que as linhas e as colunas representam o intervalo de nível de cinza e o valor  $P(l, c)$ , correspondente ao elemento da matriz na posição  $(l, c)$ , representa a frequência de níveis de cinza  $l$  e  $c$  considerando uma distância e uma direção.

Na Figura 3 é representada um exemplo de uma imagem constituída de três níveis de cinza e a matriz de co-ocorrência correspondente, sendo consideradas uma direção de  $0^\circ$  e uma distância de 1. Na posição  $(0, 1)$  da matriz de co-ocorrência encontra-se o valor  $P(0, 1) = 8$ . Esse valor é obtido considerando-se que para cada elemento da matriz da imagem onde o valor é 0, se o valor do elemento adjacente à esquerda ou a direita for igual a 1, o valor  $P(0, 1)$  da matriz de co-ocorrência é incrementado em uma unidade. Em [7] foi proposto um conjunto de 11 atributos de textura baseados em valores calculados a partir dessas matrizes de co-ocorrência: Segundo Momento Angular — SMA —, Con-

traste, Correlação, Variância, Entropia, Momento Inverso da Diferença — MID —, Soma da Média, Soma da Variância, Soma da Entropia, Diferença da Variância e Diferença da Entropia.

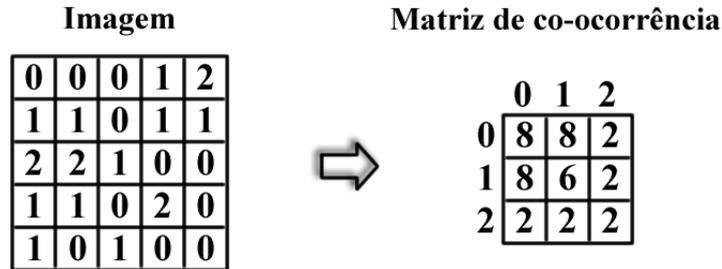


Figura 3. Imagem representada em três níveis de cinza como uma matriz de pixels 5x5 e a respectiva matriz de co-ocorrência para a direção 0° e a distância 1 [2].

### 2.3. Etapa 3: Seleção de Características Importantes

Em IA, a Seleção de Atributos é uma área de pesquisa na qual são desenvolvidos métodos para selecionar atributos considerando um determinado critério de importância, permitindo, por exemplo, reduzir a dimensão do espaço de busca e eliminar dados contendo ruídos. É desejável que os algoritmos de IA trabalhem de modo eficiente a partir de conjuntos de dados com atributos não importantes, porém, o desempenho desses algoritmos e a qualidade dos modelos resultantes, são afetados quando o conjunto de dados contém atributos irrelevantes e/ou redundantes [8].

Com isso, uma vez construído o *CDI*, devem ser selecionadas as características importantes do conjunto de dados, já que pode existir um subconjunto de atributos  $V' \subset V$  que represente bem o conjunto de exemplos. Considerando-se alguma medida de importância são obtidos os atributos que representam melhor o conjunto de dados *CDI*. Uma vez realizada a seleção de atributos importantes, cada imagem será representada por um vetor de características  $V' \subset V$  definido como  $V' = \{C'_1, C'_2, \dots, C'_{d'}\}$ , onde  $d'$  corresponde à dimensão do novo vetor  $V'$ . Desse modo, cada exemplo é descrito pelo conjunto  $E'_i = \{A'_{i,1}, A'_{i,2}, \dots, A'_{i,d'+1}\}$  composto por  $d' + 1$  atributos, o último atributo indicando a classe. Com base em *CDI* é construído o Conjunto de Dados Relevante — *CDR* — definido como  $CDR = \{E'_1, E'_2, \dots, E'_{2n}\}$ , em que cada exemplo é descrito por  $d' + 1$  atributos.

### 2.4. Etapa 4: Construção de Modelos para Classificação

Após a seleção 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 [9]. 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 *CDR*, que contém  $2n$  exemplos descritos unicamente pelos atributos que se apresentaram importantes segundo a medida de importância utilizada e pela sua classe, é realizada a construção de modelos a partir desses dados.

## 2.5. Etapa 5: Avaliação de Modelos

Na quinta 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.

## 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 [10], que foi desenvolvido em linguagem de programação JAVA<sup>1</sup> no ambiente de desenvolvimento Eclipse<sup>2</sup> e utilizando a biblioteca *Java Advanced Imaging* — JAI — [11]. O aplicativo permite apoiar tanto nos processos de seleção, operação, manipulação e visualização de imagens, como também no de 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.

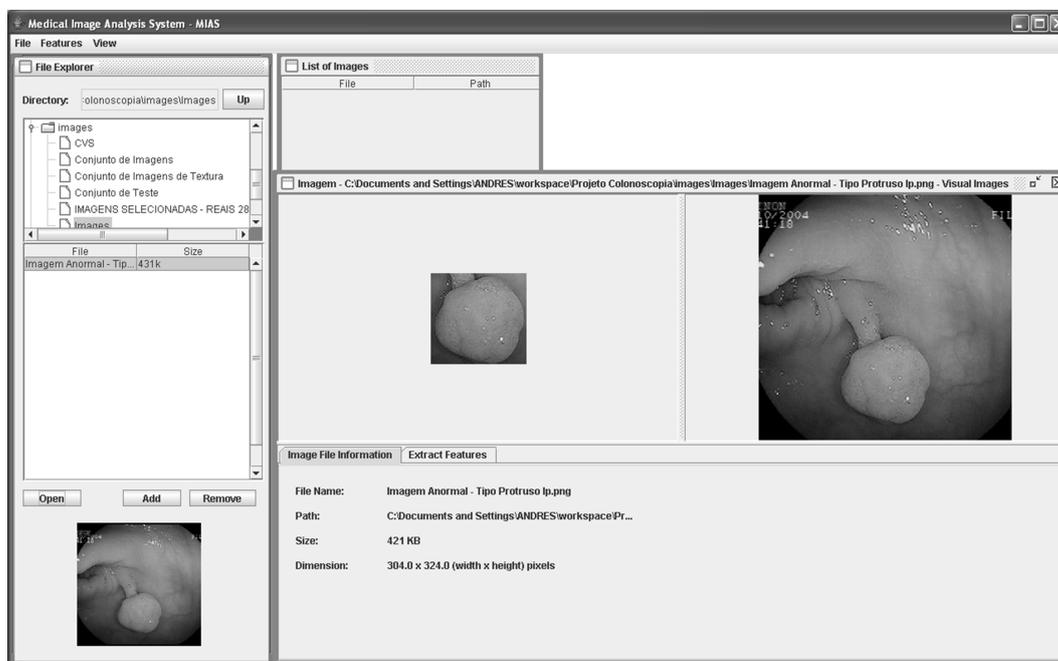


Figura 4. Screenshot da tela do aplicativo MIAS.

<sup>1</sup><http://www.java.sun.com/>.

<sup>2</sup><http://www.eclipse.org/>.

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

A metodologia apresentada na Seção 2, foi aplicada a um conjunto de 15 imagens de exames de colonoscopia contendo pólipos, coletadas do Serviço de Coloproctologia da FCM/UNICAMP. Com auxílio dos especialistas dessa instituição, inicialmente, para cada uma das imagens, nas quais não foi aplicado nenhum pré-processamento, 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  $CI = \{Im_1, Im_2, \dots, Im_{30}\}$  foi construído.

Na segunda etapa, para a representação das imagens foram utilizadas 11 características, definidas pelo conjunto  $V = \{C_1, C_2, C_3, \dots, C_{11}\}$ . Essas características — Segundo Momento Angular, Contraste, Correlação, Variância, Entropia, Momento Inverso da Diferença, Soma da Média, Soma da Entropia, Soma da Variância, Diferença da Entropia e Diferença da Variância — foram propostas em [7] e apresentadas na Seção 2.2.1. No cálculo dessas características foram consideradas matrizes de co-ocorrência para direções 0, 45, 90 e 135 graus e distância um, sendo que, para cada característica apenas a média dos valores das quatro direções foi considerado. Para a representação das imagens foi utilizado o módulo de extração de características do aplicativo MIAS. Como parâmetros específicos para a extração de características foi considerada a adjacência real (distância igual a um) de pixels para a construção das matrizes de co-ocorrência. Para isso, as imagens também 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, 12]. Desse modo, foi constituído o conjunto de dados inicial, definido por  $CDI = \{E_1, E_2, \dots, E_{30}\}$ , em que cada elemento  $E_i$  é representado por 11 características e um atributo que identifica a classe a que pertence.

A etapa de seleção de atributos foi executada utilizando a medida distância como critério de importância. Nessa seleção, foram aplicados testes estatísticos para avaliar a existência de diferença estatisticamente significativa entre os dados de cada atributo em relação às classes considerando uma significância de 95%. Para cada atributo, foi realizado o teste estatístico paramétrico  $t$  para a comparação das amostras não emparelhadas de cada classe. Para os casos em que as amostras de alguma das duas classes não apresentaram uma distribuição normal, foi utilizado o teste estatístico não paramétrico Mann-Whitney para a comparação das amostras não emparelhadas de cada classe. O aplicativo GraphPad InStat<sup>3</sup> foi utilizado para a realização dos testes estatísticos [13]. Os resultados são apresentados na Tabela 1, onde DP representa o desvio padrão, TN representa o resultado do teste de normalidade e DES a existência de diferença estatisticamente significativa.

Com base nos resultados apresentados, pode ser observado que as características Variância e Diferença da Entropia apresentaram diferença estatisticamente significativa entre os conjuntos de dados das duas classes. Com isso, a partir de CDI, foi construído o conjunto de dados relevante definido por  $CDR = \{E'_1, E'_2, \dots, E'_{30}\}$ , em que cada exemplo  $E'_i$  foi representado pelas duas características que se apresentaram importantes e pela classe a que pertence.

<sup>3</sup><http://www.graphpad.com/>.

**Tabela 1. Resultado da seleção de atributos importantes.**

Atributo	Classe com pólipó			Classe sem pólipó			<i>p-valor</i>	DES
	Média	DP	TN	Média	DP	TN		
SMA	0,1862	0,2882	Sim	0,1543	0,2858	Não	0,2169	Não
Contraste	0,6455	0,07277	Sim	0,7686	0,3205	Sim	0,2739	Não
Correlação	0,2968	0,2845	Sim	0,4218	0,3274	Sim	0,2742	Não
Variância	0,1754	0,2032	Sim	0,5223	0,2995	Sim	0,0011	Sim
Entropia	0,3765	0,2738	Sim	0,3399	0,2974	Sim	0,7287	Não
MID	0,4703	0,2311	Sim	0,2876	0,2989	Sim	0,0723	Não
Soma da Média	0,6736	0,2902	Sim	0,7087	0,3075	Não	0,5949	Não
Soma da Entropia	0,6375	0,3156	Sim	0,7383	0,2897	Sim	0,3703	Não
Soma da Variância	0,3765	0,2738	Sim	0,3399	0,2974	Sim	0,7287	Não
Diferença da Variância	0,5211	0,1895	Sim	0,712	0,2272	Sim	0,0189	Sim
Diferença da Entropia	0,3351	0,228	Sim	0,2246	0,2545	Sim	0,2212	Não

Posteriormente, foram construídos modelos para os conjuntos de dados CDI e CDR. Esses modelos foram induzidos utilizando Aprendizado Simbólico Supervisionado, por meio da geração de árvores de decisão, utilizando o algoritmo J48, o qual é uma implementação do algoritmo C4.5 [14] disponibilizado na ferramenta WEKA<sup>4</sup> [9].

Os modelos construídos foram analisados do ponto de vista da precisão preditiva e da complexidade sintática dos modelos construídos. A partir da análise das árvores de decisão geradas, pode ser observado que o número de nós e a altura da árvore foram os mesmos, evidenciando o mesmo grau de complexidade da estrutura da árvore. Outra questão interessante está relacionada aos atributos escolhidos durante a SA, os quais foram todos utilizados pelo algoritmo de indução para a construção das árvores a partir de CDI e CDR, constatando que, esses atributos, também foram considerados relevantes pelo algoritmo de indução. Para obter uma estimativa do erro médio verdadeiro (precisão) foi utilizado o método de Validação Cruzada dividindo o conjunto de dados em cinco partições. Nesse método são utilizadas, comumente, dez partições, porém devido ao número reduzido de exemplos, optou-se por utilizar uma quantidade menor de partições. Na Tabela 2 são apresentados o erro médio e o desvio padrão de classificação dos modelos construídos a partir de CDI e CDR.

**Tabela 2. Resultado do método de validação cruzada.**

	CDI	CDR
<b>Erro Médio</b>	0,3000	0,1680
<b>Desvio Padrão</b>	0,1375	0,1167

A partir dessas estimativas, pode ser observada uma diminuição de 13,3% e 2,08% nos valores de erro médio e desvio padrão, respectivamente, do modelo de CDI para o de CDR. Com base nos valores dos erros dos modelos construídos pelos conjuntos CDI e CDR, os quais passaram no teste da normalidade, foi aplicado o teste estatístico *t* para constatar a existência de diferença estatisticamente significativa entre as médias dos erros. O *p-valor* obtido foi de 0,1457, indicando a não existência de diferença estatisticamente significativa entre os erros de classificação dos modelos construídos. Portanto, a qualidade desses modelos foi estatisticamente a mesma.

<sup>4</sup><http://www.cs.waikato.ac.nz/ml/weka/>.

Para avaliar os modelos construídos foram ainda utilizadas as medidas relacionadas à tabela de contingência. Essas medidas são: Valor Preditivo Positivo — VPP —, Valor Preditivo Negativo — VPN —, Sensibilidade e Especificidade. Desse modo, na Tabela 3 e na Tabela 4 são apresentadas as tabelas de contingência dos modelos de CDI e CDR, respectivamente e na Tabela 5 são mostrados os valores das medidas anteriormente citadas.

**Tabela 3. Tabela de contingência do modelo construído a partir de CDI.**

Classificação	Tecido		Totais
	Pólipo	Sem Pólipo	
<b>Pólipo</b>	10	4	<b>14</b>
<b>Sem Pólipo</b>	5	11	<b>16</b>
<b>Totais</b>	<b>15</b>	<b>15</b>	<b>30</b>

**Tabela 4. Tabela de contingência do modelo construído a partir de CDR.**

Classificação	Tecido		Totais
	Pólipo	Sem Pólipo	
<b>Pólipo</b>	10	3	<b>13</b>
<b>Sem Pólipo</b>	5	12	<b>17</b>
<b>Totais</b>	<b>15</b>	<b>15</b>	<b>30</b>

**Tabela 5. Medidas calculadas a partir da TC dos modelos construídos.**

Modelo	VPP (%)	VPN (%)	Sensibilidade (%)	Especificidade (%)
<b>CDI</b>	71,43	68,75	66,67	73,33
<b>CDR</b>	76,92	70,59	66,67	80,00

Os dados apresentados indicam que o modelo construído a partir de CDR apresentou as seguintes características em relação ao de CDI:

- A precisão em prever a presença de pólipo foi aumentada 5,49%;
- A precisão em prever a ausência de pólipo em 1,48%;
- A probabilidade de que uma imagem de tecido cólico que contenha pólipo seja identificada com presença de pólipo foi a mesma para ambos os modelos;
- A probabilidade de que uma imagem de tecido cólico sem pólipo seja identificada com ausência de pólipo foi aumentada em 6,67%.

## 5. Conclusão

Os resultados deste trabalho mostraram que, do ponto de vista de precisão, a qualidade dos modelos construídos a partir de CDI e CDR foi estatisticamente a mesma, como também as complexidades estruturais de ambas árvores de decisão. Porém, uma análise através da TC, mostrou que o modelo construído a partir de CDR teve maior probabilidade de predição, para três das quatro medidas avaliadas, que o modelo de CDI. Neste estudo inicial, 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 esses atributos.

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 métodos para selecionar atributos relevantes considerando também o problema da redundância.

## 5. Referências

- [1] S. Karkanis, K. Galoussi, and D. Maroulis. Classification of endoscopic images based on texture spectrum. *Advance Course in Artificial Intelligence, In Proceedings of Workshop on Machine Learning in Medical Applications*, pages 63–69, 1999.
- [2] J. C. Felipe, A. J. M. Traina, and C. Traina. Retrieval by content of medical images using texture for tissue identification. *16th IEEE Symposium on Computer-based Medical Systems*, pages 26–27, 2003.
- [3] F. A. Quilici. *Colonoscopia*. Lemos-Editorial, São Paulo, Brasil, 2000.
- [4] R. S. Cotran, V. Kumar, and T. Collins. *Patologia Estrutural e Funcional*. Guanabara Koogan, USA, 2000.
- [5] S. M. Krishnan, C. J. Yap, K. V. Asari, and P. M. Y. Goh. Neural network based approaches for the classification of colonoscopic images. *Engineering in Medicine and Biology Society, Proceedings of the 20th Annual International Conference of the IEEE*, 3:1678–1680, 1998.
- [6] P. M. D. A. Marques, M. H. Honda, J. A. H. Rodrigues, R. R. Santos, A. J. M. Traina, and C. Traina. Recuperação de imagens baseada em conteúdo: Uso de atributos de textura para caracterização de microcalcificações mamográficas. *Revista Brasileira de Radiologia*, 35:93–98, 2002.
- [7] R. Haralick, K. Shanmugam, and I. Dinstein. Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics*, 3(6):610–621, 1973.
- [8] H. D. Lee. *Seleção de atributos importantes para a extração de conhecimento de bases de dados*. PhD thesis, Instituto de Ciências Matemáticas e de Computação (ICMC), Universidade de São Paulo (USP), São Paulo - Brasil, 2005.
- [9] I. H. Witten and E. Frank. *Data mining: practical machine learning tools and techniques*. Elsevier, 2 edition, 2005.
- [10] C. A. Ferrero, H. D. Lee, F. C. Wu, C. S. R. Coy, J. J. Fagundes, and J. R. N. Góes. Seleção de características baseadas em textura para a identificação de anormalidades em imagens de colonoscopia. *X Congresso Brasileiro de Informática em Saúde*, pages 1–6, 2006.
- [11] L. H. Rodrigues. *Building imaging applications with java technology*. Addison Wesley, Boston, USA, 2001.
- [12] P. Howarth, A. Yavlinsky, D. Heesch, and S. Rüger. Medical image retrieval using texture, locality and colour. *Lecture Notes from the Cross Language Evaluation Forum*, pages 740–749, 2005.
- [13] H. Motulsky. *Intuitive Biostatistics*. Oxford University Press, New York, USA, 1995.
- [14] J. R. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco, USA, 1993.