



**FORMULÁRIO DE PROJETO DE TRABALHO
PROINOVA, ProBIC, PIBIC, Artigo 170 e PIPG**

1. TÍTULO DO PROJETO DE TRABALHO:

Descritor de Características para Classificação e Localização da Expressão de Proteínas em Imagens Patológicas de Imuno-Histoquímica

2. ÁREA DE CONHECIMENTO (Conforme CNPq - por extenso)

2.1 Sub-área de conhecimento (Conforme CNPq - por extenso)

1.03.03.05-7 Processamento Gráfico (Graphics)

2.2 Grupo de Pesquisa e respectiva área:

Grupo de Inteligência Aplicada

3. RESUMO:

Na área médica, tipicamente o diagnóstico de câncer é feito pelo patologista através de uma inspeção visual da morfologia e cromaticidade de amostras de tecidos. Uma atividade laboriosa onde a efetividade do método depende diretamente da atenção e conhecimento do especialista. O desenvolvimento contínuo em tecnologias relacionadas à bioimagem, especialmente na microscopia, vem estabelecendo técnicas guiadas por computador para avaliação de imagens biológicas como uma forma efetiva para extração de informações clínicas e funcionais a partir de amostras de tecido. A Imuno-histoquímica permite o monitoramento e a quantificação da expressão de proteínas com fins de diagnóstico, prognóstico, planejamento de tratamento e pesquisa. Os anticorpos são marcados com uma coloração específica, o que viabiliza a aplicação de técnicas de visão computacional para auxiliar no processo de classificação. Com este foco, este projeto tem como objetivo geral definir um descritor de características que combine informações de cromaticidade de um dos espaços de cores com o método para descrição de textura conhecido por Padrão Binário Local. A partir deste objetivo será possível aplicar esses descritores no contexto da Imuno-histoquímica e auxiliar no processo de classificação de patologias. Para que esse objetivo seja atingido, será aplicado uma metodologia que se enquadra como hipotético-dedutivo, bem como uma abordagem quantitativa quando se propõem a executar experimentações e avaliações empregando métodos estatísticos.

Palavras-chave: 1: Visão computacional; 2: Descritor de características; 3: Imuno-histoquímica.

4. INTRODUÇÃO:

A Análise de Imagens Médicas (AIM) vem crescendo nos últimos 40 anos, inicialmente o tema era considerado uma sub-área que relacionava disciplinas estabelecidas como processamento de imagens, reconhecimento de padrões e visão computacional, até se tornar uma área de pesquisa distinta. Fato este que proporcionou o surgimento de uma comunidade interdisciplinar envolvendo áreas como ciência da computação, engenharias, matemática, biologia e medicina. Uma consulta ao PubMed¹ em 30 de março de 2016 buscando o termo AIM, retornou um total de 19.342 trabalhos entre 1976-1995, já entre 1996-2015 retornou um total de 161.948 trabalhos. Esses números mostram o interesse e a expansão do campo pela comunidade científica (FRANGI, TAYLOR, GOOYA, 2016).

A patologia digital, um campo dentro da AIM, vem ganhado interesse clínico e científico devido ao avanço no desenvolvimento de tecnologias de microscopia e imageamento digital, bem como no aumento da capacidade de processamento dos computadores. A digitalização de lâminas de vidro com amostras de tecido biológico tem facilitado a telepatologia, que quando aliada a tecnologias de telecomunicações vem possibilitado a transmissão de imagens em alta resolução, com dados patológicos onde a finalidade pode ser relacionada a pesquisa, diagnóstico ou educação (MADABHUSHI, LEE, 2016; HAMILTON et al., 2014).

Na área médica, tipicamente o diagnóstico de câncer é feito pelo patologista através de uma inspeção visual da morfologia e cromaticidade de amostras de tecidos. Uma atividade laboriosa e sujeita a erros (intra e extra observador), onde a efetividade do método depende diretamente da atenção e conhecimento do especialista. O desenvolvimento contínuo em tecnologias relacionadas à bioimagem, especialmente na microscopia, vem estabelecendo técnicas guiadas por computador para avaliação de imagens biológicas como uma forma efetiva para extração de informações clínicas e funcionais a partir de amostras de tecido (MOUELHI et al., 2013; CATALDO, FICARRA, MACII, 2012).

O desenvolvimento de tecnologias voltadas a sistemas para diagnóstico médico guiado por computador tornou-se um dos principais temas de pesquisa em imagens médicas e radiodiagnóstico, uma vez que cresce o consenso entre especialistas da área de análise de imagens e patologia sobre a importância da análise quantitativa de imagens patológicas. Essas análises envolvem tanto o ponto de vista do diagnóstico clínico, quanto à área de pesquisa, proporcionando o entendimento dos mecanismos biológicos das doenças (SAMSI et al., 2010).

A Imuno-histoquímica (IHQ) é uma tecnologia aplicada no diagnóstico patológico desde a década de setenta, amplamente utilizada como uma ferramenta para a detecção de proteínas em amostras de tecido. A IHC permite o monitoramento e a quantificação da expressão de proteínas com fins de diagnóstico, prognóstico, planejamento de tratamento e

¹ <https://www.ncbi.nlm.nih.gov/pubmed/>

pesquisa. Os anticorpos são marcados com uma coloração específica, e a intensidade da referida expressão está relacionada com a intensidade da coloração alcançada no processo (LIU, WANG, FANG, 2006; WEMMERT et al., 2013; FICARRA et al., 2006) (vide figura 01).

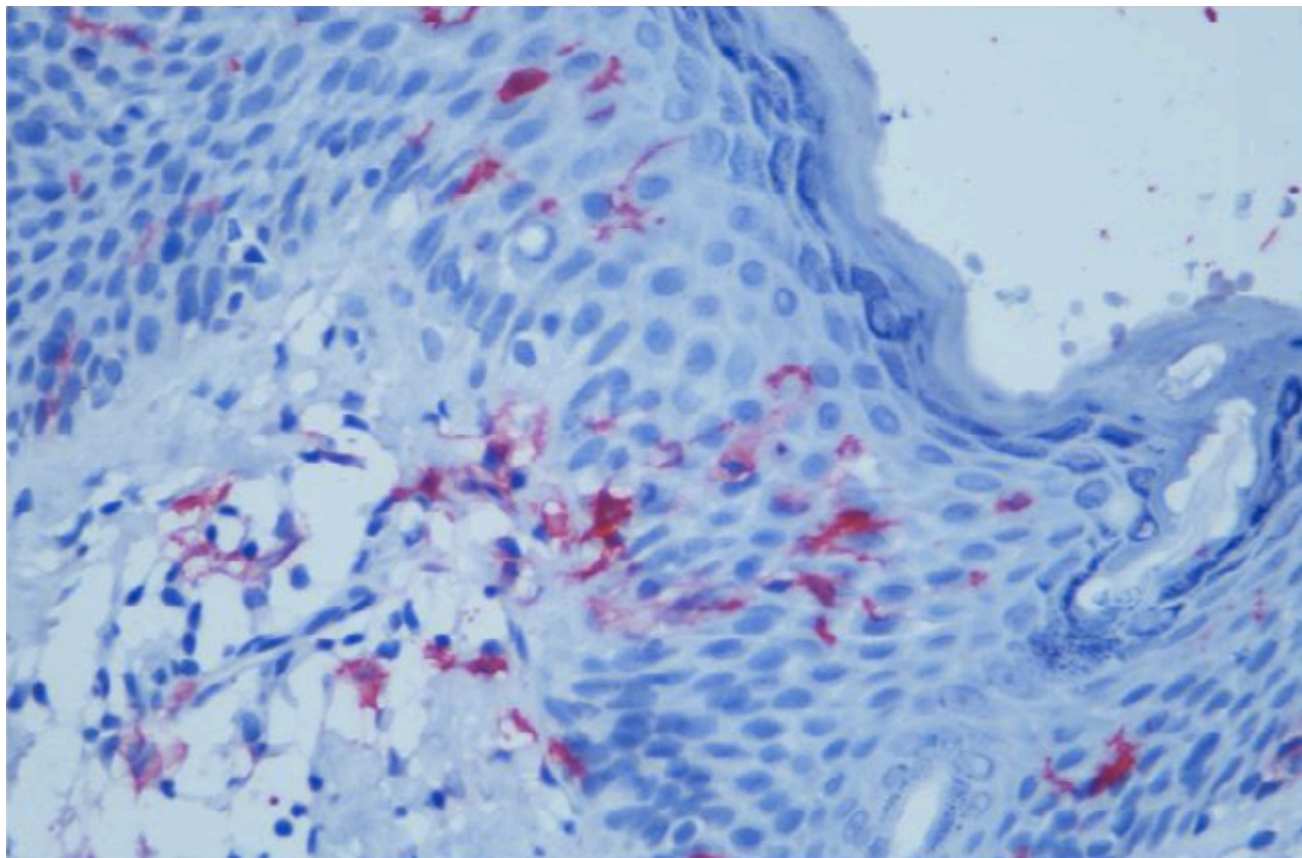


Figura 01: Imagem de Imuno-histoquímica

Embasado nessas questões levantadas, na experiência e nas capacidades de desenvolvimento de tecnologias na área de visão computacional instaladas no 4Vision Lab, esse se projeto apresenta os elementos necessários para o desenvolvimento deste projeto de pesquisas.

5. PROBLEMA:

Normalmente as proposta encontradas na literatura para a classificação e localização da expressão de proteínas em imagens de IHC são desenvolvidas para um tecido ou marcador de coloração específicos. Diante do contexto apresentado até aqui, formulou-se a seguinte pergunta de pesquisa: é possível combinar atributos de cor e textura para representar padrões de IHC, que possam ser utilizados com qualquer tecido ou marcador de cromaticidade?

6. OBJETIVOS:

6.1. Objetivos gerais:

Definir um descritor de características que combine informações de cromaticidade de um dos espaços de cores com o método para descrição de textura conhecido por Padrão Binário Local.

6.2. Objetivos específicos:

- Revisão bibliográfica sobre abordagens de AIM voltadas a Imuno-histoquímica;
- Pesquisar e definir o descritor a ser utilizado no contexto deste projeto;
- Modelar do protótipo de software a ser desenvolvido;
- Implementar um protótipo de software;
- Avaliar o desempenho do descritor em uma base de imagens estabelecida.

7. JUSTIFICATIVA:

Conforme previamente apresentado, a Imuno-histoquímica (IHQ) é uma tecnologia aplicada no diagnóstico patológico, amplamente utilizada como uma ferramenta para a detecção de proteínas em amostras de tecido. A IHC Permite o monitoramento e a quantificação da expressão de proteínas com fins de diagnóstico, prognóstico, planejamento de tratamento e pesquisa. Os anticorpos são marcados com uma coloração específica, e a intensidade da referida expressão está relacionada com a intensidade da coloração alcançada no processo. Com base nesse processo, torna-se viável aplicar diversas tecnologias de processamento digital e imagens afim de auxiliar no processo de diagnóstico.

Desenvolver descritores de características mais precisos podem tornar esse processo de auxílio mais efetivo e seguro para os especialistas da área médica. A partir desse descritor, estimasse que muitas novas pesquisas e aplicações possam ser beneficiadas, o que justifica o desenvolvimento deste projeto de pesquisas.

8. REVISÃO BIBLIOGRÁFICA/FUNDAMENTAÇÃO TEÓRICA

Nesta seção são apresentados os fundamentos teóricos necessários para a elaboração e execução deste projeto. Dentre os temas abordados encontra-se o métodos extratores de características, algoritmos de segmentação de Imagens e detectores de objetos. Esses temas fazem parte dos algoritmos e fases básicas para a pipeline de processamento digital de imagens.

8.1 Extratores de características

Em si tratando de vetores de características, eles são representados por letras maiúsculas em negrito, como x , onde cada componente x_i , representa a i -ésima medida que compõem o descritor. Normalmente os vetores são expressos na forma de matrizes $n \times 1$, a natureza de cada medida que compõem um vetor de características x depende da metodologia empregada para descrever o padrão físico propriamente dito.

No trabalho proposto por Sobieranski é apresentada uma abordagem para identificação de coloração imuno-histoquímica em imagens microscópicas de amostra de pele. Baseado em um processo de aprendizado definido como Distância Polinomial de Mahalanobis (DPM), que emprega a projeção polinomial da Distância de Mahalanobis (DM) para a construção de uma função de similaridade.

O método DPM permite a classificação de padrões não lineares de IHQ em imagens microscópicas de histopatologia. A abordagem referida emprega atributos de cor no modelo cromático Vermelho, Verde e Azul (red, green e blue – RGB), como Descritores de Características (DC) para o treinamento da métrica de discriminação. Relacionados, seguindo os critérios: (i) combinar informações de cor e textura como DC; e (ii) utilizar um método de classificação supervisionada.

A fusão de imagens consiste no processo de compor uma única imagem a partir de um conjunto de imagens de uma mesma observação, a imagem resultante apresenta maior informação. Um processo amplamente utilizado na formação de imagens, com aplicação em áreas com vigilância, visão de máquina, sensoriamento remoto e imageamento médico.

O algoritmo LBP consiste em um método popular na descrição de texturas encontrado na literatura, onde um descritor invariante, baseado em níveis de cinza é composto através da comparação de um píxel central com seus píxeis vizinhos. O método LBP demonstra ser um descritor de baixa complexidade computacional e menos sensível a variações de luminosidade do que muitos descritores, sendo utilizado em muitas tarefas de classificação com resultados promissores, ainda conforme (LIU, 2017) é um dos métodos para descrição de textura mais proeminentes e bastante explorado pela comunidade científica.

A Figura 2 mostra uma visão geral da formação do DC. As medidas estatísticas escolhidas para compor o DC inicial serão a média, o desvio padrão, a assimetria, a curtose e a energia. Tais medidas serão tomadas em cinco vizinhanças distintas com dimensões de 3x3, 5x5, 7x7 e 9x9 (na imagem resultante do filtro de textura LBP), a fim de descrever um único píxel. Desta forma o DC inicial será formado por três valores de cor e valores de textura.

Na Figura 2 a imagem no canto superior esquerdo representa a imagem original de IHC, imediatamente a direita desta, a imagem com fundo preto representa o resultado da aplicação do método LBP. Na metade inferior da Figura 1 está representado a construção do DC, que será formado por uma componente de cor (de acordo com o espaço de cor a ser selecionado), tomada da imagem original. Bem como uma componente com informações de textura, tomada da imagem filtrada. Desta forma cada píxel será descrito por um vetor de características, composto a partir da imagem original e da imagem resultante do processamento de filtragem com método LBP.

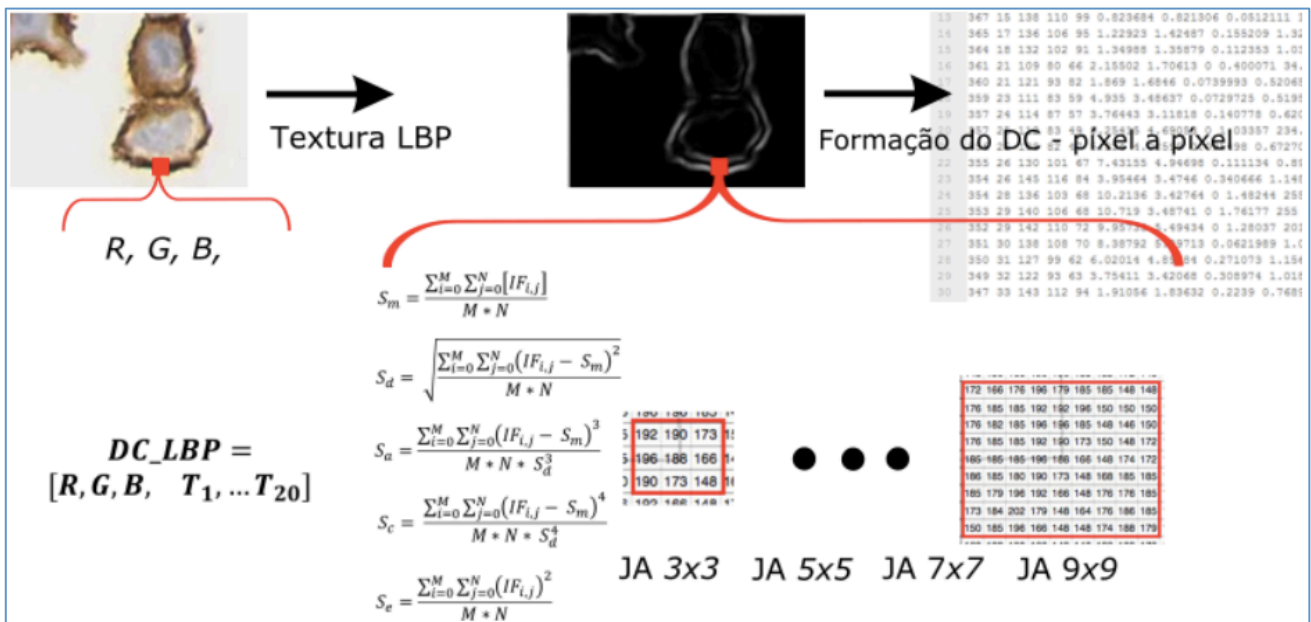


Figura 02: Visão geral da formação do DC, a partir da imagem original e da imagem filtrada será construído um descritor 20-dimensional para representar um píxel

8.2 Segmentação de Imagens

A segmentação particiona o conjunto de dados de entrada em estruturas com um conteúdo que seja relevante para a aplicação em que está inserida. Essas estruturas de saída são objetos ou partes de objetos que auxiliarão o processo de interpretação da imagem (PEDRINI; SCHWARTZ, 2008).

Abordagens convencionais de segmentação de imagem normalmente são baseadas nas propriedades básicas dos níveis de cinza da imagem, buscando detectar descontinuidade ou similaridade na imagem. A descontinuidade visa particionar a imagem com base em mudanças abruptas nos níveis de cinza, caracterizando assim a presença de pontos isolados, linhas ou bordas na imagem. Já as similaridades procuram agrupar pontos da imagem que apresentam valores similares para um determinado conjunto de características. A seleção dessas características está relacionada ao domínio da aplicação (PEDRINI; SCHWARTZ, 2008).

Atualmente, a maioria dos métodos de segmentação usam informações temporal ou espacial das imagens (WANG et al., 2003). Algumas abordagens convencionais serão descritas a seguir.

8.3 Detecção de Objetos

De acordo com Pratt (2007), uma das maneiras mais fundamentais de detecção de objetos dentro de um campo de imagem é o casamento de modelos (do inglês *template matching*), em que uma réplica do objeto de interesse é relacionada a todos os objetos desconhecidos

no campo da imagem. Se a relação entre o modelo e o objeto desconhecido for suficientemente próxima, o objeto desconhecido é rotulado como objeto modelo.

Raramente um modelo será exatamente igual ao objeto de interesse, e uma incerteza *à priori* é a forma exata e a estrutura do objeto a ser detectado. Consequentemente, um procedimento comum é produzir uma medida de diferença $D(m, n)$, entre o modelo $T(m, n)$, e todos os pontos do campo da imagem $F(j, k)$. Um objeto é considerado correspondente, sempre que a diferença é menor do que um nível estabelecido $L_R(m, n)$. A medida de diferença usual é a diferença média quadrada ou erro, tal como definido pela Equação 1 a seguir: (PRATT, 2007).

$$D(m, n) = \sum_j \sum_k [F(j, k) - T(j - m, k - n)]^2 \quad \text{Equação 1}$$

Com o objetivo de considerar outros dois elementos, energia e a correlação cruzada $R_{FT}(m, n)$, entre o campo da imagem e o modelo, a Equação 1 é expandida e passa a ser formalizada pela Equação 2, e a existência do modelo no campo da imagem em comparação é estabelecida de acordo com a Equação 3.

$$\tilde{R}_{FT}(m, n) = \frac{\sum_j \sum_k [F(j, k)T(j - m, k - n)]}{\sum_j \sum_k [F(j, k)]^2} \quad \text{Equação 2}$$

$$\tilde{R}_{FT}(m, n) > L_R(m, n) \quad \text{Equação 3}$$

Outras fontes literárias apresentam o termo *template matching* como *block matching*, além de outras fórmulas e critérios para definição da relação entre o modelo, ou bloco, e m relação a imagem de interesse.

9. METODOLOGIA:

Formalmente a pesquisa a ser conduzida neste trabalho irá se enquadrar no método hipotético-dedutivo, pois irá combinar pesquisa bibliográfica e avaliação experimental para formar conhecimento. O trabalho irá assumir as características, partindo da sua natureza, como uma pesquisa aplicada pelo fato de tratar um problema específico relacionado à AIM de imuno-histoquímica. Assume também uma abordagem quantitativa quando se propõem a executar experimentações e avaliações empregando métodos estatísticos, estabelecidos na literatura.

Segundo seus objetivos, o trabalho pode ser classificado como uma pesquisa exploratória uma vez que propõem o levantamento bibliográfico, apresentando ainda características de uma pesquisa descritiva, uma vez que irá avaliar a relação de causa e efeito entre as variáveis do processo, tais como parâmetros dos algoritmos envolvidos, componentes do DC e desempenho de classificação (taxas de acerto e erros). Consequentemente, os procedimentos técnicos adotados para concretizar o trabalho proposto foram a pesquisa bibliográfica e a pesquisa experimental.

Todo o conjunto de dados selecionado para avaliação neste trabalho será extraído do Human Protein Atlas (HPA) (UHLEN, 2015). As imagens utilizadas serão relativas ao gene

CD44, a proteína codificada por este gene é classificada como um biomarcador para candidato a câncer. Estas imagens proveem de 56 linhas de células distintas, de origens variadas (p. ex. cérebro, fígado, mama, pele) e são relacionadas ao anticorpo CAB000316. Nesta categoria linha de célula do HPA, são fornecidas imagens de sobreposição (metadados), com anotação pixel a pixel avaliadas por especialistas, com informações de intensidade e localização da expressão de proteínas.

Três grupos de experimentos serão desenvolvidos para avaliar o DC definido: (i) Dimensionalidade, que consistiu em avaliar o grau de importância que cada dimensão do DC tem em relação a representatividade do padrão de interesse; (ii) Desempenho, que consistirá em executar uma série de testes avaliando o resultado de classificação; e (iii) Variância que irá usar uma análise de variância (teste d hipótese) entre os desempenhos de classificação (utilizando o descritor com e sem textura).

10. CRONOGRAMA DE ATIVIDADES DE PESQUISA

A1. Pesquisa Bibliográfica, será contemplada a execução da revisão sistemática da literatura sobre métodos de descrição, classificação e métricas de avaliação, relacionadas a análise de imagens de imuno-histoquímica;

A2. Levantamento bibliográfico sobre métodos para descrição de textura;

A3. Definição do Descritor, será avaliado a contribuição de cada componente de textura, empregando uma análise de componentes principais, com a finalidade de reduzir a dimensionalidade do DC inicial (20-dimensional):

A4. Será executado o levantamento de requisitos, bem como a criação dos diagramas de modelagem para guiar o processo de desenvolvimento do protótipo;

A5. Implementação do protótipo, com a finalidade de avaliar o descritor;

A6. Experimentos, foi planejada para avaliar desde a análise de dimensionalidade do descritor definido inicialmente, passando por experimentos para avaliar o desempenho de classificação, combinado o descritor com três abordagens distintas de classificação.

10.1 Cronograma físico da pesquisa

[illegible]

11. REFERÊNCIAS

- CATALDO, S. D., FICARRA, E., MACII, E. (2012). Computer-aided techniques for chromogenic immunohistochemistry: Status and directions. *Computers in Biology and Medicine*, 42(10):1012 – 1025.
- FICARRA, E., MACII, E., DE MICHELI, G., AND BENINI, L. (2006). Computer-aided evaluation of protein expression in pathological tissue images. In *Proceedings of the 19th IEEE Symposium on Computer-Based Medical Systems, CBMS '06*, pages 413–418, Washington, DC, USA. IEEE Computer Society.
- FRANGI, A. F., TAYLOR, Z. A., GOOYA, A. (2016). Precision imaging: more descriptive, predictive and integrative imaging. *Medical Image Analysis*, 33:27 – 32.
- HAMILTON, P. W., BANKHEAD, P., WANG, Y., HUTCHINSON, R., KIERAN, D., MCART, D. G., JAMES, J., SALTO-TELLEZ, M. (2014). Digital pathology and image analysis in tissue biomarker research. *Methods*, 70(1):59 – 73. Advancing the boundaries of molecular cellular pathology.
- LIU, B., WANG, W., FANG, X. (2006). Automatic extraction of positive cells in tumor immunohistochemical pathology image based on ycbcr. In *2006 6th World Congress on Intelligent Control and Automation*, volume 2, pages 9708–9712.
- LIU, L., FIEGUTH, P., GUO, Y., WANG, X., PIETIKAINEN, M. (2017). Local binary features for texture classification: Taxonomy and experimental study. *Pattern Recognition*, 62:135 – 160.
- MADABHUSHI, A. LEE, G. (2016). Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical Image Analysis*, 33:170 – 175. 20th anniversary of the *Medical Image Analysis* journal (MedIA).
- MOUELHI, A., SAYADI, M., FNAIECH, F., MRAD, K., ROMDHANE, K. B. (2013). Automatic image segmentation of nuclear stained breast tissue sections using color active contour model and an improved watershed method. *Biomedical Signal Processing and Control*, 8(5):421 – 436.
- PEDRINI, H.; SCHWARTZ, W.R., *Análise de Imagens Digitais: Princípios, Algoritmos, e Aplicações*. São Paulo: Thomson Learning. 2008.
- PRATT, W.K. *Digital Image Processing*. New Jersey: John Wiley & Sons, Inc. 2007.
- WANG, L.; HU, W.; TAN, T. Recent developments in human motion analysis. *Pattern Recognition*, 36 (3), 2003.
- SAMSI, S., LOZANSKI, G., SHANARAH, A., KRISHANMURTHY, A. K., GURCAN, M. N. (2010). Detection of follicles from ihc-stained slides of follicular lymphoma using iterative watershed. *IEEE Transactions on Biomedical Engineering*, 57(10):2609–2612.
- UHLÉN, M., FAGERBERG, L., HALLSTROM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, A., KAMPF, C., SJO`STEDT, E., ASPLUND, A., OLSSON, I., EDLUND, K., LUNDBERG, E., NAVANI, S., SZIGYARTO, C. A.-K., ODEBERG, J., DJUREINOVIC, D., TAKANEN, J. O., HOBER, S., ALM, T., EDQVIST, P.-H., BERLING, H., TEGEL, H., MULDER, J., ROCKBERG, J., NILSSON, P., SCHWENK, J. M., HAMSTEN, M., VON FEILITZEN, K., FORSBERG, M., PERSSON, L., JOHANSSON, F., ZWAHLEN,

M., VON HEIJNE, G., NIELSEN, J., PONTEN, F. (2015). Tissue-based map of the human proteome. *Science*, 347(6220).

WEMMERT, C., KRÜGER, J. M., FORESTIER, G., STERNBERGER, L., FEUERHAKE, F., GANCARSKI, P. (2013). Stain unmixing in brightfield multiplexed immunohistochemistry. In 2013 IEEE International Conference on Image Processing, pages 1125–1129.