

Arcabouço para Classificação, Recuperação por Conteúdo e Radiômica de Imagens Médicas: uma investigação de biomarcadores quantitativos para o câncer de pulmão

José Raniery Ferreira Junior^{1,2}, Marcel Koenigkam Santos²,
Paulo Mazzoncini de Azevedo Marques²

¹Escola de Engenharia de São Carlos (EESC), Universidade de São Paulo (USP)
CEP 13566-590 – São Carlos-SP – Brasil

²Faculdade de Medicina de Ribeirão Preto (FMRP), Universidade de São Paulo (USP)
CEP 14040-900 – Ribeirão Preto-SP – Brasil

jose.raniery@alumni.usp.br, {marcelk46, pmarques}@fmrp.usp.br

Abstract. *This work aimed to decrease the feature (inefficient numeric representation of images) and performance (low quality of integration) gaps that clinical decision support systems may have. Some medical image datasets were produced during the development of this research. Semiautomatic models of pattern recognition, content-based retrieval, image classification, and prognosis evaluation were developed to build a framework for medical image analysis. The framework disclosed the potential to integrate with health information systems and identified several biomarkers, potentially supporting decision making. Finally, this thesis contributed to the identification of a radiomics signature to assess lung cancer diagnosis and prognosis for patient risk stratification.*

Resumo. *Este trabalho tentou diminuir as limitações de característica (representação numérica ineficiente das imagens) e de desempenho (qualidade baixa de integração) que sistemas de suporte a decisões clínicas podem apresentar. Alguns conjuntos de imagens foram produzidos para a pesquisa. Modelos semiautomáticos de reconhecimento de padrões, recuperação por conteúdo, classificação e avaliação prognóstica foram desenvolvidos para compor um arcabouço de análise de imagens médicas. O arcabouço mostrou potencial de integração com sistemas de informação em saúde e identificou diversos biomarcadores. Por fim, esta tese contribuiu para a identificação de uma assinatura radiômica do câncer de pulmão para a estratificação de risco do paciente.*

1. Caracterização do problema e motivação da pesquisa

Sistemas de auxílio computadorizado ao diagnóstico (CAD - *Computer-Aided Diagnosis*) têm sido desenvolvidos para auxiliar especialistas na detecção precoce de doenças [Gillies et al. 2016]. O objetivo do CAD é melhorar a acurácia e consistência da interpretação e diagnóstico por imagem, mediante o uso da sugestão de resposta (segunda opinião) fornecida por um computador [Ferreira and Oliveira 2017]. Após a revolução tecnológica na área de diagnóstico por imagem com a estruturação de ambientes radiológicos totalmente digitalizados, houve a integração destes ambientes com outros sistemas de informação em saúde. O principal exemplo desta integração é o sistema de

arquivamento e comunicação de imagens médicas (PACS - *Picture Archiving and Communication System*), responsável por receber, armazenar e disponibilizar as imagens no padrão de comunicação de imagem digital em medicina (DICOM - *Digital Imaging and Communications in Medicine*) dos diversos dispositivos de aquisição [Santos et al. 2019].

A integração do PACS com os sistemas de informação clínica possibilitou o surgimento de um outro modelo de auxílio computadorizado ao diagnóstico, baseado na recuperação de imagens visualmente similares. Com o objetivo de aumentar a aplicabilidade clínica do PACS, foram incorporadas técnicas de recuperação de imagens baseada em conteúdo (CBIR - *Content-Based Image Retrieval*) nos sistemas para a radiologia [Ferreira et al. 2017]. A CBIR tem sido descrita como uma das ferramentas computacionais mais promissoras, pois ela pode auxiliar o processo de decisão clínica, recuperando em grandes bases de dados casos similares já previamente diagnosticados [Ferreira and Oliveira 2015]. Tradicionalmente, sistemas CAD executam tarefas de classificação de imagens, fornecendo uma única resposta ao especialista, e sistemas CBIR executam buscas de imagens médicas por similaridade, fornecendo um conjunto de casos similares a um caso desconhecido dado como entrada. Porém, nenhum dos dois tipos de tecnologia fornece diretamente informações de auxílio ao prognóstico e à decisão terapêutica, o que limita o escopo da sua aplicabilidade à rotina clínica.

Em contrapartida, a radiômica tem sido descrita como uma extensão do CAD que correlaciona atributos extraídos de imagens médicas obtidas em exames de rotina com desfechos clínicos do paciente, aumentando o poder dos modelos de suporte à decisão [Wu et al. 2016, Aerts et al. 2014]. Em face dos recentes avanços em terapias alvo para a medicina de precisão, a necessidade de uma abordagem simples de análise robusta de imagens médicas se tornou imperativa, e a radiômica pode prover isto por ser uma ferramenta não-invasiva, rápida e de baixo-custo para a rotina clínica [Ferreira et al. 2020, Sacconi et al. 2017]. A análise radiômica define de forma explícita um processo de extração massiva de atributos das imagens médicas, de inserção destes atributos em bases que agreguem outros dados e informações do paciente que possam ser compartilhadas, e que permitam a geração e/ou a confirmação de hipóteses diagnósticas, prognósticas e de tratamento [Gillies et al. 2016]. A integração dos dados clínicos é ainda mais importante à medicina de precisão devido o potencial preditivo da mineração dos dados do paciente na atual “Era *Big Data*”. Neste contexto, a radiômica pode ser entendida como a translação dos conceitos de *Big Data* e medicina de precisão para o ecossistema do diagnóstico por imagem e está focada no desenvolvimento de ferramentas e bases de apoio à tomada de decisão clínica, que possam potencialmente melhorar o diagnóstico, o prognóstico e a precisão de uma previsão de resposta terapêutica [Santos et al. 2019].

Como principal justificativa desta pesquisa, é citada a ausência de modelos e biomarcadores robustos de classificação, recuperação e radiômica de imagens médicas que permitam o auxílio computadorizado ao diagnóstico, prognóstico e decisão terapêutica. O desenvolvimento de um arcabouço integrado poderia aumentar a aplicabilidade dos sistemas de informação em radiologia e contribuir para a melhora da performance diagnóstica, da avaliação do prognóstico do paciente e acompanhamento da progressão da doença. Este trabalho ainda tentou diminuir duas lacunas apresentadas por modelos de apoio à decisão clínica: de característica, referente à representação numérica eficiente das imagens; e de desempenho, referente à implementação e qualidade na integração,

avaliação e generalização/replicação dos modelos.

2. Objetivos e contribuições do trabalho

O objetivo principal desta tese foi desenvolver um arcabouço para classificação, recuperação por conteúdo e radiômica de imagens médicas que possa ser aplicado no apoio às decisões clínicas. O câncer de pulmão em imagens de Tomografia Computadorizada (TC) foi definido como caso de uso para o desenvolvimento da pesquisa, devido a diversos desafios, como por exemplo, a classificação dos tumores nas imagens de TC e a avaliação do prognóstico do paciente.

Este trabalho investigou diferentes métodos de CAD, CBIR e radiômica para o desenvolvimento do arcabouço para a análise de imagens médicas, visando contribuir para o avanço tecnológico na área radiológica. Um modelo semiautomático de reconhecimento de padrões foi desenvolvido baseado em segmentação por crescimento de região 3D, extração de atributos *fine-tuned*, classificação e recuperação por conteúdo otimizada por GPU (*Graphics Processing Unit*) de imagens de TC e avaliação prognóstica do paciente. Um segundo modelo semiautomático baseado em aprendizado profundo agregado e redes neurais convolucionais também foi proposto com o objetivo de diminuir as etapas de pré-processamento e extração de atributos do arcabouço para a análise computadorizada das imagens. Os modelos apresentaram potencial para compor uma ferramenta possível de ser integrada na rotina clínica. A integração com sistemas hospitalares poderia permitir o encaminhamento mais rápido do paciente com prioridade elevada para a realização dos exames complementares de rotina e tratamento.

3. Resultados obtidos e discussão

No primeiro momento da pesquisa, foi realizado o desenvolvimento de um algoritmo de análise de nitidez de borda de imagens de TC armazenadas em uma infraestrutura computacional não-relacional baseada em nuvem [Ferreira and Oliveira 2015, Ferreira et al. 2016]. Além disso, foi realizada a classificação de nódulos pulmonares em termos de potencial malignidade baseada em atributos de matriz de coocorrência de níveis de cinza e de nitidez de borda. Os atributos combinados com um classificador de floresta aleatória apresentaram a maior performance no reconhecimento de padrões, gerando resultados promissores para o diagnóstico auxiliado por computador. Entretanto, uma árvore de decisão com apenas dois atributos (a amplitude dos níveis de cinza de linhas perpendiculares ao nódulo e o momento de diferença inverso (MDI) calculado da matriz de coocorrência) foi uma solução de baixo custo computacional para uma ferramenta CAD, com performance estatisticamente equivalente a obtida pela floresta aleatória com todos os atributos para a classificação das imagens [Ferreira Junior et al. 2018].

O propósito do desenvolvimento de um algoritmo de CBIR neste trabalho foi complementar o reconhecimento de padrões das imagens baseado na busca por casos similares de uma imagem de referência que o especialista pode ter dificuldades de interpretar ou sobre a qual esteja buscando por mais informações clínicas. Apesar da performance (em termos de velocidade) não ser um fator crítico no diagnóstico do câncer de pulmão, foi optada a otimização com GPU para não acarretar no desinteresse do usuário em utilizar o algoritmo de CBIR com baixo desempenho em velocidade [Ferreira et al. 2017, Ferreira and Oliveira 2017]. Dois atributos estatísticos de nitidez

de borda propostos na tese (amplitude e curtose) e quatro de textura de matriz de co-ocorrência (contraste, proeminência, variância e MDI) apresentaram a maior eficiência na busca por similaridade de imagens de TC e para o diagnóstico diferencial de nódulos pulmonares. A recuperação baseada nestes seis atributos de TC e otimização em GPU representou uma solução de baixo custo computacional e boa eficiência para uma ferramenta de auxílio ao diagnóstico do câncer de pulmão [Ferreira Junior et al. 2017].

As análises radiômicas realizadas nesta pesquisa identificaram diversos atributos com potencial diagnóstico e prognóstico, associados com metástase linfonodal, metástase distante, histopatologia e avaliação prognóstica do paciente com câncer de pulmão. Reconhece-se que um radiologista com a leitura de um exame de TC de tórax poderia obter uma performance maior na detecção de metástase linfonodal do que a obtida neste trabalho. Porém a performance do especialista poderia ser melhorada com a inclusão de uma ferramenta de radiômica na rotina clínica. Além disso, a avaliação de metástase distante é tradicionalmente realizada em tumores secundários espalhados em sítios distantes do paciente que precisariam de outros exames de imagem, como TC ou ressonância magnética de cabeça, cintilografia de osso, ou ainda ultrassonografia ou TC abdominal. Existem poucas evidências da presença de metástases a partir de apenas análises do sítio primário [Zhou et al. 2018, Coroller et al. 2015]. Logo, uma abordagem que proporcione mais informações sobre a provável existência de metástases distantes baseada na avaliação do tumor primário em um exame de TC de tórax realizado no início da assistência médica ao paciente, como a proposta neste trabalho, poderia reduzir consideravelmente o tempo para a adoção de um exame de imagem complementar, podendo resultar em uma decisão mais rápida para a terapia adequada ao paciente. Ademais, a radiômica da histopatologia poderia fornecer mais informações clínicas importantes que potencialmente aumentariam a confiança do radiologista na predição do resultado de uma biópsia ou ressecção cirúrgica. Por fim, a avaliação prognóstica do paciente pela metodologia aqui proposta poderia proporcionar um avanço aos modelos clínicos atuais baseados em informações radiológicas subjetivas e exames genômicos invasivos por ser objetiva, não-interpretativa, não-invasiva e de baixo custo.

Este trabalho realizou novas descobertas para a radiômica do câncer de pulmão em comparação com o que foi encontrado na literatura. Por exemplo, os atributos de energia de *wavelet*, medidas de informação de correlação e de ocupação de matrizes de níveis de cinza nunca haviam sido anteriormente associados com metástases até o momento da escrita da tese [Ferreira et al. 2018]. A direcionalidade da textura representada por atributos de Tamura e a estimativa da dimensão fractal estão associadas com a histopatologia e ambas as metástases, e isto também nunca não havia sido anteriormente relatado na literatura [Ferreira et al. 2020]. Atributos estatísticos extraídos no domínio espectral de Fourier nunca foram investigados antes na associação com a sobrevida global dos pacientes e a heterogeneidade intratumoral [Ferreira Jr et al. 2020]. Além disso, nenhum trabalho foi encontrado analisando padrões de metástase distante no tumor primário com uma abordagem de aprendizado profundo agregado [Ferreira et al. 2019a], nem analisando características morfológicas específicas dos tumores (i.e., cavidade e pseudocavidade) com uma abordagem radiômica [Ferreira et al. 2019b]. Por fim, esta tese também confirmou alguns achados anteriormente descritos pela comunidade acadêmica, como o atributo de volume de forma [Coroller et al. 2015] e o gênero do paciente [Zhou et al. 2018, Coroller et al. 2015] estarem associados com a presença de metástase

distante e o atributo de probabilidade máxima da COM [Zhu et al. 2018], o diâmetro clínico e o gênero do paciente [Digumarthy et al. 2019] com o subtipo tumoral encontrado no exame histopatológico.

Por fim, o trabalho foi capaz de construir modelos computacionais que poderiam trazer avanços aos sistemas de informação em saúde, mostrou o potencial de diferentes métodos tecnológicos na avaliação quantitativa de imagens médicas e ainda a possibilidade de substituição de técnicas tradicionais por métodos modernos de inteligência artificial, resultando em *pipelines* com menor número de processos. Além disso, é enfatizado o potencial da CBIR paralela ao CAD e à radiômica para o suporte clínico com a busca e fornecimento aos especialistas de informações clínicas obtidas de casos visualmente similares. Este trabalho também mostrou a possibilidade e importância do ajuste fino nos diferentes processos de análise quantitativa de imagens, como de extração de atributos e aprendizado de máquina. Do ponto de vista radiológico, o arcabouço pode proporcionar diversos biomarcadores quantitativos de imagens para diferentes doenças, em conjunto com informações clínicas importantes para os especialistas, potencialmente auxiliando as tomadas de decisões da rotina clínico-radiológica. Por fim, esta tese ainda contribuiu para a descoberta de uma assinatura radiômica de avaliação diagnóstica e prognóstica para a estratificação de risco de vida dos pacientes com câncer de pulmão.

Agradecimentos

Esse trabalho foi financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Código de Financiamento 001), Programa de Doutorado Sanduíche no Exterior (PDSE-CAPES, processo nº 88881.134004/2016-01), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, processo nº 452257/2018-2) e Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, processo nº 2016/17078-0).

Referências

- Aerts, H., Velazquez, E., Leijenaar, R., Parmar, C., Grossmann, P., Carvalho, S., Bussink, J., Kains, B., Rietveld, D., Hoebers, F., Rietbergen, M., Leemans, C., Dekker, A., Gillies, R., and Lambin, P. (2014). Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nature Communications*, 5:4006.
- Coroller, T., Grossmann, P., Hou, Y., Velazquez, E., Leijenaar, R., Hermann, G., Lambin, P., Kains, B., Mak, R., and Aerts, H. (2015). CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol*, 114(3):345–350.
- Digumarthy, S., Padole, A., Gullo, R., Sequist, L., and Kalra, M. (2019). Can CT radiomic analysis in nslc predict histology and EGFR mutation status? *Medicine*, 98:e13963.
- Ferreira, J., Santos, M., Cipriano, F., Fabro, A., and Marques, P. (2018). Radiomics-based features for pattern recognition of lung cancer histopathology and metastases. *Computer Methods and Programs in Biomedicine*, 159:23–30.
- Ferreira, J., Santos, M., Hironaka, T., Cipriano, F., Fabro, A., Yoshida, H., and Marques, P. (2019a). Deep learning-based radiomics of primary lung tumors in CT images for prediction of distant metastasis. *Sup Int J Comput Assist Radiol Surg*, page S79.
- Ferreira, J., Santos, M., Kikuchi, Y., Faleiros, M., Cipriano, F., Fabro, A., Yoshida, H., and Marques, P. (2019b). A radiomics approach for differentiation of pseudocavitation from cavitation on lung cancer tumors. *Sup Int J Comput Assist Radiol Surg*, page S78.

- Ferreira, J., Santos, M., Tenório, A., Faleiros, M., Cipriano, F., Fabro, A., Näppi, J., Yoshida, H., and Marques, P. (2020). CT-based radiomics for prediction of histologic subtype and metastatic disease in primary malignant lung neoplasms. *Int J Comput Assist Radiol Surg*, 15:163–172.
- Ferreira, J. R., Oliveira, M. C., and Azevedo-Marques, P. M. (2016). Cloud-based NoSQL open database of pulmonary nodules for computer-aided lung cancer diagnosis and reproducible research. *Journal of Digital Imaging*, 29(6):716–729.
- Ferreira, J. R., Oliveira, M. C., and Azevedo-Marques, P. M. (2017). Integrating 3D image descriptors of margin sharpness and texture on a GPU-optimized similar pulmonary nodule retrieval engine. *The Journal of Supercomputing*, 73(8):3451–3467.
- Ferreira, Junior, J. R. and Oliveira, M. C. (2015). Evaluating margin sharpness analysis on similar pulmonary nodule retrieval. In *Proceedings of 28th IEEE International Symposium on Computer-Based Medical Systems*, pages 60–65.
- Ferreira, Junior, J. R. and Oliveira, M. C. (2017). Unidade de processamento gráfico no diagnóstico auxiliado por computador. *Journal of Health Informatics*, 9(3):103–108.
- Ferreira Jr, J., Santos, M., Machado, C., Faleiros, M., Correia, N., Cipriano, F., Fabro, A., and Marques, P. (2020). Análise radiômica do câncer de pulmão para avaliação prognóstica do paciente e da heterogeneidade intratumoral. *Radiol Bras*, Ahead.
- Ferreira Junior, J. R., Azevedo-Marques, P. M., and Oliveira, M. C. (2017). Selecting relevant 3D image features of margin sharpness and texture for lung nodule retrieval. *Int J Comput Assist Radiol Surg*, 12(3):509–517.
- Ferreira Junior, J. R., Oliveira, M. C., and Azevedo-Marques, P. M. (2018). Characterization of pulmonary nodules based on features of margin sharpness and texture. *Journal of Digital Imaging*, 31(4):451–463.
- Gillies, R., Kinahan, P., and Hricak, H. (2016). Radiomics: Images are more than pictures, they are data. *Radiology*, 278(2):563–577.
- Sacconi, B., Anzidei, M., Leonardi, A., Boni, F., Saba, L., Scipione, R., Anile, M., Rengo, M., Longo, F., Bezzi, M., Venuta, F., Napoli, A., Laghi, A., and Catalano, C. (2017). Analysis of ct features and quantitative texture analysis in patients with lung adenocarcinoma: a correlation with egfr mutations and survival rates. *Clin Radiol*, 72:443–50.
- Santos, M., Ferreira, J., Wada, D., Tenório, A., Barbosa, M., and Marques, P. (2019). Artificial intelligence, machine learning, computer-aided diagnosis, and radiomics: advances in imaging towards to precision medicine. *Radiologia Brasileira*, 52(6):387–396.
- Wu, J., Shultz, D., Gudur, M., Rubin, D., Loo, B., Diehn, M., and Li, R. (2016). Early-stage non-small cell lung cancer: Quantitative imaging characteristics of 18f fluorodeoxyglucose pet/ct allow prediction of distant metastasis. *Radiology*, 281:270–78.
- Zhou, H., Dong, D., Chen, B., Fang, M., Cheng, Y., Gan, Y., Zhang, R., Zhang, L., Zang, Y., Liu, Z., et al. (2018). Diagnosis of distant metastasis of lung cancer: Based on clinical and radiomic features. *Translational Oncology*, 11(1):31–36.
- Zhu, X., Dong, D., Chen, Z., Fang, M., Zhang, L., Song, J., Yu, D., Zang, Y., Liu, Z., Shi, J., et al. (2018). Radiomic signature as a diagnostic factor for histologic subtype classification of non-small cell lung cancer. *European Radiology*, 28(7):2772–2778.