Osteosarcoma is the most common form of bone cancers, with 5-year survival rate of about 60% [1]. In this issue of EBioMedicine, Wu et al. developed a computational method to predict personalized 5 years-survival status from Computed Tomography (CT) scans of high grade osteosarcoma imaged at the time of diagnosis [11]. By extracting hundreds of numbers (features) from each CT scan, they trained a machine-learning predictive model and validated its performance using information from 150 patients imaged in their institution. This study has followed the footsteps of previous studies that used similar approaches. For example, to predict lymph node metastasis in colorectal cancer [2], prognosis-associated biomarkers in non–small cell lung cancer [3], and survival in non–small cell lung cancer [4]. Such computational-based conversion of medical images to high-dimensional quantitative representations, and their use to identify patterns in such datasets is termed Radiomics, and has great potential to impact cancer detection, diagnosis, treatment and prognosis [5].

Many challenges face toward making Radiomics more effective in the clinic: change of culture, bureaucracy matters and methodology. Here, I briefly discuss several opportunities to improve the technical/methodological aspects that can drive forward the field.

First, it is important to recognize the existence of a community of computer scientists and engineers that apply computer vision to medical imaging. Researchers in this community have their own venues to present and publish their research (e.g., MICCAI, ISBI) and peer-assessment is mostly aimed toward algorithmic novelty and surpassing state-of-the-art performance with some, but not sufficient, focus on the clinical relevance. Bridging these gaps between the clinicians and computational scientists is key in advancing the field. Accordingly, efforts for synergetic collaborative projects are becoming more common in recent years (for example, the MICCAI challenges [6] and http://miccai.cloudapp.net/competitions/84).

Deep learning has emerged as a promising method to extract information from images [7]. Its success stems from the idea of optimizing the feature-representation of an image – moving away from manually-engineered to automatic- and application-specific feature extraction. Such approach requires “big data” to train robust models that can generalize well to capture the essence of the data, an aspect that is mostly lacking from patient-derived imaging data where collection of large datasets is complicated and expensive. This limitation can be resolved with transfer learning, modifying existing models, trained for another purpose, by retraining them with the specific data of interest [8]. This approach already proven success in the realm of medical imaging, even when the initial model was trained on non-medical images [9].

Another promising opportunity lies in data integration across different sources, containing complementary information, to enhance performance and predictive accuracy. Wu et al. demonstrated that combining CT-derived features with clinical factors outperforms models trained using each of these sources of information separately [11]. This conclusion was previously confirmed by others, including Grossmann et al. [10] who convincingly demonstrated integration of radiomics (CT), genetic, and clinical information. Advanced computational methods can be exceedingly beneficial to extract more predictive information by data-fusion of imaging modalities, genomics and clinical data.

Making medical images publicly available is critical for rapid technological advancements. This approach has proven success for omics-data, where the field of bioinformatics emerged on the grounds of large-scale collection of data, openly disseminated according to the FAIR principles - findable, accessible, interoperable, and reusable. This is even more complicated for medical imaging due to privacy issues, large scale, plethora of imaging technologies, data formats, and variability in imaging protocols between clinicians and institutions. Metadata harmonization and organization is key for effective data deposition, dissemination and retrieval, and can be achieved through multi-disciplinary community-driven standardization efforts between clinicians, academics, industrial and regulatory representatives. Large repositories must be built to host large-scale medical image data from multiple institutions. To fully exploit the clinical potential, the image data should be accompanied by the corresponding clinical information, diagnosis, prognosis, and eventual outcome, when available. Funding this expensive infrastructure should be via public funding agencies and private philanthropy. The availability of such massive datasets is essential to fully exploit the information encapsulated in medical images, and
indeed, efforts in this direction are just beginning to sprout (e.g., http://www.cancerimagingarchive.net/, https://aimi.stanford.edu/medical-imagenet). It will be important to assess the variability and batch-effects within and between different imaging modalities, protocols and institutions as a first step toward data harmonization, integration and knowledge-retrieval.

Radiomics has the potential to significantly improve diagnosis and personalized care of cancer patients. Bridging the gap between the clinical and computational community toward synergetic efforts is necessary to tackle the technical challenges and will eventually lead to improved care routinely practiced in the clinic.

The author declares no conflicts of interest.

Acknowledgments

I would like to thank Idit Diamant and Noga Levy for discussions and constructive feedback on the manuscript.

References

[1] Mirabello L, Troisi RJ, Savage SA. Osteosarcoma incidence and survival rates from 1973 to 2004. Cancer 2009;115:1531–43.

[2] Huang Y-q, Liang C-h, He L, Tian J, Liang C-s, Chen X, et al. Development and validation of a radiomics nomogram for preoperative prediction of lymph node metastasis in colorectal cancer. J Clin Oncol 2016;34:2157–64.

[3] Gevaert O, Xu J, Hoang CD, Leung AN, Xu Y, Quon A, et al. Non–small cell lung cancer: identifying prognostic imaging biomarkers by leveraging public gene expression microarray data—methods and preliminary results. Radiology 2012;264:387–96.

[4] Huang Y, Liu Z, He L, Chen X, Pan D, Ma Z, et al. Radiomics signature: a potential biomarker for the prediction of disease-free survival in early-stage (i or ii) non–small cell lung cancer. Radiology 2016;281:947–57.

[5] O’connor JP, Aboagye EO, Adams JE, Aerts HJ, Barrington SF, Beer AJ, et al. Imaging biomarker roadmap for cancer studies. Nat Rev Clin Oncol 2017;14:168.

[6] Abdallah K, Hugh-Jones C, Norman T, Friend S, Stolovitzky G. The prostate cancer DREAM challenge: A community-wide effort to use open clinical trial data for the quantitative prediction of outcomes in metastatic prostate cancer. AlphaMed Press; 2015.

[7] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems; 2012. p. 1097–105.

[8] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Advances in neural information processing systems; 2014. p. 3120–8.

[9] Van Ginneken B, Setio AA, Jacobs C, Ciompi F. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. Bio medical Imaging (ISBI), 2015 IEEE 12th International Symposium on; 2015. p. 286–9.

[10] Grossmann P, Stringfield O, El-Hachem N, Bui MM, Velazquez ER, Parmar C, et al. Defining the biological basis of radiomic phenotypes in lung cancer. Elife 2017;6:e27421.

[11] Wu Y. Survival Prediction in High-grade Osteosarcoma Using Radiomics of Diagnostic Computed Tomography. EBioMedicine 2018;34:27–34.