Deep Medical Image Computing in Preventive and Precision Medicine

Le Lu
Ping An Technology
US Research Labs
Johns Hopkins University

Adam P. Harrison
NVIDIA

Deep learning has a game-changing potential to improve the state of preventative and precision medicine within medical image computing. Here, we will first overview preventative and precision medicine and field of deep learning. Afterward, we will share our perspective on recent research and development activities in both areas and point out some existing achievements, positive indications, limitations, and near future opportunities and impediments. To flesh out our viewpoints, we draw from examples of our most recent work, which largely stem from radiologic images, but we encourage readers to consult some other recent reviews, which include many references that space did not allow us to include. We also assume the reader is broadly familiar with machine learning technologies.

Overview and Status: *Preventive medicine* in medical imaging refers to early detection of disease findings, e.g., lung nodules, colonic polyps, and liver/bone lesions, with the goal of timely patient intervention and management. Traditionally, this is done using manual examination from noninvasive imaging modalities, but more recently computer-aided solutions are becoming more prominent. *Precision medicine*, within imaging, means computing quantitative and precise imaging biomarkers, e.g., volumetric tumor measurements for tracking and beyond, to support clinical decision making and ultimately improve patient outcomes. Current radiological practices are still largely qualitative, even for the most advanced medical centers.

Deep learning, especially deep convolutional neural networks (CNNs), have made significant headway in categorically improving both preventative and precision medicine. This includes the preventative detection of anomalous findings in various imaging modalities, such as histology images or within computed tomography (CT) scans. For instance, markedly higher quantitative performance has been achieved for classifying enlarged lymph nodes and colonic polyps from CT.[2,4] For precision medicine, progress has been made on accurately segmenting organs[1,3,6] and anatomical anomalies,[7] which would play a central role for any quantitative markers.

The main reason of these early successes is *that effective learning of hand-crafted features for medical image analysis problems is notoriously hard, whereas CNNs eliminate this need*. In light of this, for the first time, CNNs have made feasible large-scale medical image parsing and tagging (over thousands or tens of thousands of patients and studies).[4,9,11] Works using CNNs have also built up a massive body of empirical evidence indicating that low-level features can be shared and fine-tuned

 Published by the IEEE Computer Society

between networks trained on different image modalities or even from networks trained on natural image understanding tasks, e.g., from the ImageNet competition.

Deep learning has also rekindled and intensified industrial interest in medical imaging applications. Currently, there is a healthy body of startups focusing on medical image analysis and informatics, e.g., HeartFlow, Enlitic, Arterys, Viz.ai, Zebra-Me, and Butterfly Network. These efforts complement the research and development initiatives from large corporations, such as Siemens, IBM, Tencent, and Google Brain/DeepMind. Industrial investment and engagement cover various topics in both preventive and precision medicine. Though many technological, business, and clinical challenges lay ahead, scalable and effective deep learning principles will continue propelling high-performance, deployable medical imaging, and clinical informatics applications for years to come.

Future Directions in Deep Preventive Imaging Medicine: Early developments in computer-aided detection (CADe), in the predeep learning era, mainly concentrated on detecting breast lesions/masses from mammography scans and detecting colonic polyps, lung nodules, vascular lesions, and pulmonary embolisms from CT or CT angiography images. Commercial clinical software products from several vendors were developed and deployed into practices, but the "expected" broad success and uptake did not happen. CADe products typically operate in a second reader mode (the *de facto* protocol approved by FDA), which means a radiologist first finishes an independent image reading without CADe and then examines the CADe generated findings to make his or her final decision. This pipeline is designed to increase detection sensitivity, with the aid of CADe software, but at the cost of additional physician workload. A common drawback is that the above-mentioned applications are not too difficult for experienced radiologists, and the extra workload is not always compensated by the moderate to minimal increases in sensitivity, if any.

New pain points: In order for preventive CADe systems and applications really to take off, new and true clinical pain points, which are not possible to fully address yet, need to be tackled and solved. These should lie beyond the traditional second reader protocol or an extra layer of safety. Instead, they should more actively aim to improve patient care capabilities. We provide a nonexhaustive list of a few promising examples as follows.

1. First reader triage software to (potentially) significantly increase the chances of detecting, and therefore quickly treating, patients suffering from a large vessel occlusion (LVO) from a stroke has recently been cleared by the FDA (work from Viz.ai). The current manual LVO stroke workflow results in low rates and long delays of treatment, which can be alleviated by this computer-aided triage and notification software that also saves stroke specialists' time. Another similar work is atherosclerotic vascular calcification detection and segmentation using low-dose full-body CT scans, which is a very time-consuming task and easy to miss for human readers. Long-standing drawbacks of manual exams and the high performance of deep learning alternatives meant that these tasks were amenable to a CADe approach. Additional opportunities for first reader software need to be identified and seized to further expand the impact of CADe solutions.
2. Chest X-rays are the most common medical imaging exams and a very accessible modality for screening both healthy (annual health exams) and unhealthy populations, e.g., those found in community clinics and hospitals, respectively. A game-changing application would be a reliable and economical automated chest x-ray screening and referral tool deployed across massive populations, especially those that are geographically distant from major hospitals. A total of four technical challenges, not necessarily specific to chest x-rays, stand in the way of such a vision: a) chest x-rays are associated with higher degrees of diagnosis uncertainty, whether analyzed by radiologists or computerized systems, than other modalities, e.g., lung nodule detection using chest CT; b) an extremely low false positive rate is required for generic preventive screening since a large majority of a population will be healthy; c) modeling and incorporating disease ontology is critical for reasoning and regularizing the raw outcomes from image classifiers to produce sensible diagnoses; d) human interpretable and verifiable results are required to produce a clinically complete CADe system. Work is ongoing to overcome these challenges, but recent developments, e.g., weakly supervised visual grounding of disease locations,[11] bring the field closer to this vision.
3. Full body preventive organ anomaly and cancer screening is the holy grail for general and asymptomatic population screening. Most likely, the ideal setup would also incorporate

cheaper and less intrusive nonimaging technologies, such as the "CancerSeek" blood test,[8] to screen all patients undergoing annual health exams. Next, a high-performance, high-accuracy automated medical imaging organ and pathology segmentation tool could be used to localize and verify the initial finding indications. Finally, a clinical decision fusion module, combining all nonimaging and imaging test results, could report and refer the identified "high risk" patients, versus the vast majority of average-risk individuals, to specialists. Although recent work on detecting and segmenting especially difficult organs and anomalies have made good progress,[3,6] for general population preventive screening, extremely generalizable deep learning methods, possibly trained on massive datasets, require further investigation.

Future Directions in Deep Precision Imaging Medicine: Compared to preventative medical imaging, precision imaging has not been as well studied. Historically, quantitative imaging has faced roadblocks due to insufficiencies of prevailing machine learning technologies and a lack of buy-in from clinical partners in running the clinical trials and/or opening up the data archives needed to discover, characterize, and validate quantitative biomarkers. However, with the increasing capabilities of deep learning and prominent policy-level pushes for precision medicine, we see tremendous opportunity.

New pain points: Efforts should focus on computing precision imaging biomarkers at *hospital scale*, bringing forth analyses that physicians desire but are out of reach of human capabilities alone. These should focus on markers for prominent morbidities, especially cancer, but they should also provide tools that allow entirely new types of retrospective analyses for biomarker discovery. A key capability that requires further development is how to train deep learning systems on existing data sources, such as hospital archives, that are very large scale, but also messy and unstructured.

1. A common prerequisite of precision medicine is accurate and robust segmentation of anatomical structures from medical scans, i.e., classifying every pixel or voxel into a semantic meaning. Due to their superior performance, deep CNN-based segmentation methods[1,3,6] are now predominant. The value of using segmentation techniques is that raw image scans can be converted into semantic and human interpretable features, such as the volume of the left ventricle or the shape statistics of a patient's pancreas. These organ/anatomy based shape, volume, and appearance statistics can be computed from 2-D/3-D/4-D imagery, to assist both personalized diagnosis and treatment and also large population profiling. An important challenge is collecting enough data for training and ensuring any segmentation solution is generalizable to patient distributions encountered "in the wild."
2. When it comes to cancer, precision tumor growth tracking and prediction are additional key elements. Deep learning has pushed the capabilities of both forward. For instance, physicians need scalable solutions to intelligently match, track, and provide evidence-based similarity measurements to measure tumor growth rates from multiple time point studies of a patient. Due to the difficulty in obtaining training data, recent works train deep learning models on messy and large-scale clinical databases.[7,9] An example is illustrated in Figure 1. Continuing to leverage these large-scale data sources will be the key in further improving tracking capabilities. A related initiative is using deep learning techniques to observe subtle and precise longitudinal imaging changes in order to predict tumor growth rates and patterns.[10] A visual example of tumor growth prediction modeling and comparison is shown in Figure 2. Both tasks are tackling critical and clinically useful precision imaging biomarker problems, which cannot be done by human doctors alone due to the need to ingest "big data" to make accurate measurements and predictions.
3. Last but not the least, one of the ultimate goals of the precision medicine is performing retrospective analyses on clinical data to discover new imaging biomarkers that are correlated with morbidity. This can be framed as disease/concept discovery and tagging, given hospital-scale, or better yet multi-institutional, data of patient images and nonimaging records. This will likely require modeling multimodal imaging and nonimaging patient data on a graph configuration that builds and preserves pairwise[9] or higher order patient similarities. Such a representation could provide an indexable and holistic patient data view and repository, allowing analyses beyond plain classification. Importantly, given the long-tailed distribution of many diseases or ailments, such analyses are highly difficult, if not impossible, to perform without powerful computerized techniques, such as deep learning, that can effectively leverage data at massive scales.

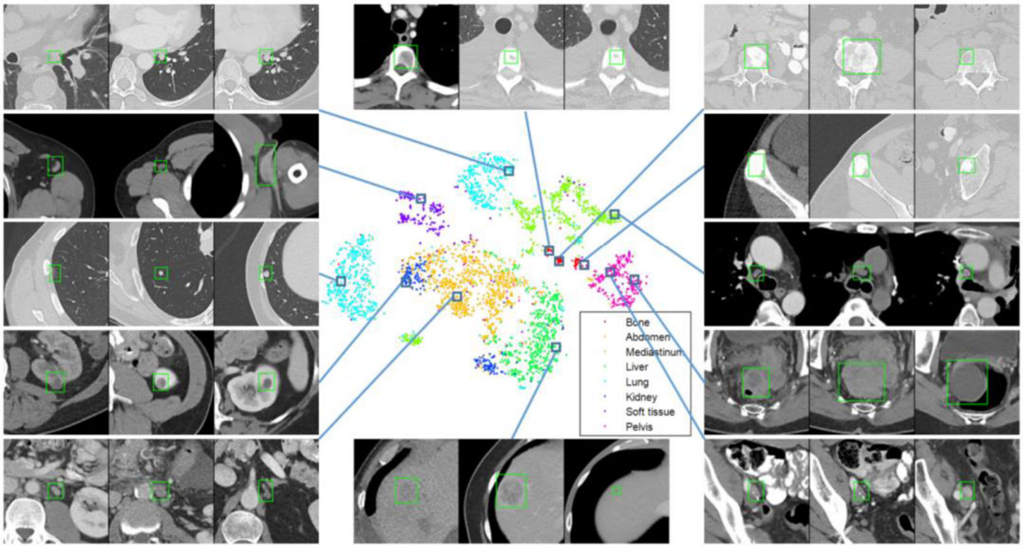

Figure 1. Visualization of clusters of tumor types automatically discovered from an analysis of a large scale dataset of ~34 000 tumors automatically extracted from a hospital archive.[9] Colors indicate the manually labeled lesion types, which correspond well with the automatically discovered tumor types. Best viewed in color.

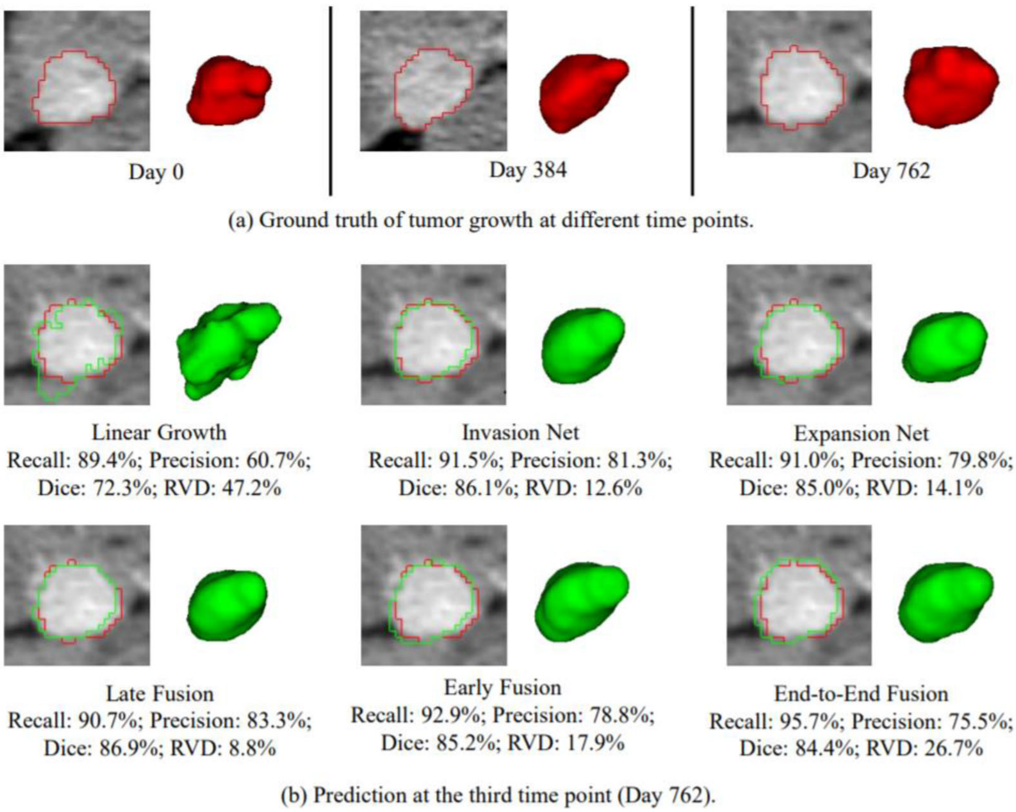

Figure 2. An example of the tumor growth prediction using a deep learning system.[10] (a) The segmented (ground truth) tumor contours and volumes at different time points. (b) The prediction results at the third time point from various automatic systems. Deep learning based tools match well with the ground truth. Red and green represent ground truth and predicted boundaries, respectively.

In summary, recent deep learning developments have been very impactful for medical imaging problems and applications, even can make some important tasks (e.g., first reader triage) from impossible via nondeep principles to reach possibly a clinical relevance level of performance. It will be a promising but challenging path going forward.

REFERENCES

1. A. P. Harrison et al., “Progressive and multi-path holistically nested neural networks for pathological lung segmentation from CT images,” in *Proc. MICCAI*, 2017, pp. 621–629.
2. H. Roth et al., “Improving computer-aided detection using convolutional neural networks and random view aggregation,” *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1170–1181, May 2016.
3. H. Roth et al., “Spatial aggregation of holistically-nested convolutional neural networks for automated pancreas localization and segmentation,” *Med. Image Anal.*, vol. 45, pp. 94–107, 2018.
4. H. Shin et al., “Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning,” *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.
5. H. Shin et al., “Interleaved text/image deep mining on a large-scale radiology database for automated image interpretation,” *J. Mach. Learn. Res.*, vol. 17, no. 107, pp. 1–31, 2016.
6. J. Cai et al., “Improving deep pancreas segmentation in CT and MRI images via recurrent neural contextual learning and direct loss function,” in *Proc. MICCAI*, 2017, pp. 674–682.
7. J. Cai et al., “Accurate weakly supervised deep lesion segmentation on CT scans: Self-paced 3D mask generation from RECIST,” 2018, arXiv: 1801.08614.
8. J. Cohen et al., “Detection and localization of surgically resectable cancers with a multi-analyte blood test,” *Science*, vol. 359, pp. 926–930, 2018.
9. K. Yan et al., “Deep lesion graphs in the wild: Relationship learning and organization of significant radiology image findings in a diverse large-scale lesion database,” in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2018.
10. L. Zhang, L. Lu, R. M. Summers, E. Kebebew, and J. Yao, “Convolutional invasion and expansion networks for tumor growth prediction,” *IEEE Trans. Med. Imag.*, vol. 37, no. 2, pp. 638–648, Feb. 2018.
11. X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “ChestX-Ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, 2017, pp. 3462–3471.
12. X. Wang et al., “Unsupervised joint mining of deep features and image labels for large-scale radiology image categorization and scene recognition,” in *Proc. IEEE Winter Conf. Appl. Comput. Vision*, 2017, pp. 998–1007.

ABOUT THE AUTHORS

Le Lu can be contacted at le.lu@pingan.com.

Adam P. Harrison can be contacted at aharrison@nvidia.com.

This article originally appeared in IEEE MultiMedia, *vol. 25, no. 3, 2018.*