Artificial intelligence in medical image analysis

Summary

Recent decades witnessed an exponential growth in the number of studies investigating the potential applications of artificial intelligence (AI) in medical image analysis. Handcrafted radiomics is one of the methods that employ AI methods in medical image analysis. Handcrafted radiomic features (HRFs) are quantitative features extracted from medical images by applying handcrafted formulas on the array of values representing a given medical image. The general hypothesis is that HRFs can decode biologic characteristics, and could be used potentially to personalize patients care. In addition, handcrafted radiomics can become an alternative to current gold standard diagnostic methods, as with proper development, it can be a non-invasive and time-saving clinical tool. HRFs have been reported to correlate with different clinical endpoints, such as classification of lesions on medical images, predicting response to therapy, and survival.

Despite the reported high potential of handcrafted radiomics, a number of limitations that hinders the clinical integration of radiomics signatures have been identified. HRFs have been reported to be sensitive to inter-reader variability, test-retest, and to variations in imaging parameters, in addition to the need for large datasets. In this thesis, we performed experiments to validate these hypotheses. We confirmed that HRFs are sensitive to the above mentioned variations, using phantom and patients reproducibility studies. We further hypothesized that different harmonization methods will have different effects on different HRFs. We performed experiments to assess the impacts of different harmonization methods, mainly image resampling and ComBat harmonization. Lastly, we hypothesized that a quantitative tool can be developed based on the differences in imaging parameters. Our novel MPenn radiomics reproducibility score was developed using a large number of scenarios of variations in imaging parameters, and has shown robustness and high performances in assessing the percentage of reproducible HRFs across scans acquired differently. The score can be utilized in future radiomics studies to evaluate the agreement in HRFs, if the scans to be analyzed are acquired differently. The score would also help interpreting the results of radiomics analyses.

Additionally, we performed a number of experiments to assess potential applications of deep learning (DL), the other AI method investigated in this thesis, in medical image analysis. Classification of bone scintigraphies and the automated detection and segmentation of non-small cell lung carcinoma on CT scans are the two tasks investigated in this thesis. For each of the tasks, multicenter data was collected, and a relatively large number of medical images were used to train the DL algorithms. A partition of the collected datasets was kept for external validation of developed algorithms. In addition, in silico trials to assess the performance of developed algorithms were designed for each of the tasks investigated. Our results showcased the potential of DL algorithms to be used as clinical decision support tools, with one of the developed algorithms receiving CE mark.

In conclusion, this thesis has confirmed a number hypothesis regarding the applications of handcrafted radiomics and deep learning in medical image analysis. For handcrafted

radiomics, we proposed and assessed a workflow for robust handcrafted radiomics analyses that will help developing generalizable radiomics signatures; and developed a novel quantitative method to assess the reproducibility of HRFs across scans acquired differently. For DL, we assessed and showcased the potential of automated algorithms to aid clinical decision making. We developed a DL algorithm for three different tasks, which showed high performance and prospective for clinical applications.