Artificial intelligence for automated radiomics for non-small cell lung cancer

Dr Jack Ross, Dr Mitch Chen, Prof Sue Copley, Prof Eric Aboagye, Prof Andrea Rockall, Imperial College London

Introduction

Aim: to develop a pipeline for automating objective and quantitative measures of lung cancer imaging for prognosis and prediction

Lung cancer exhibits phenotypic characteristics that can be captured on medical imaging. Some subtle signs that are indiscernible to the human eye can be analysed by machine learning. There is increasing evidence that these radiomic features can be used to help assess clinical outcomes including survival, histology, treatment response and adverse effects. However, these studies have been limited by the time-consuming image segmentation.

This project aimed to develop a pipeline for a fully automated deep learning-based lung cancer segmentation and radiomic feature reproducibility analysis.
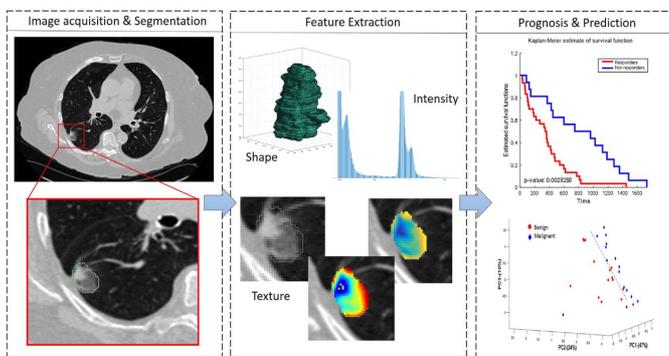
Radiomics

Radiomics is the high-throughput extraction of quantitative features from imaging to objectively and quantitatively describe tissues.

This compares to traditional subjective descriptions of lesions shape & structure ("spiculated, septated, round") as well as texture & intensity ("homogenous, uniformly enhancing"). These quantitative features are hopefully able to better predict patient outcomes such as prognosis and response to treatment. Machine learning techniques can also be used to quantify higher level features such as wavelet decomposition, that are harder for humans to identify but have been shown to correlate with outcomes.

Workflow

1. Patient identification and anonymisation
2. Tumour segmentation
3. Pre-processing, feature extraction and feature selection
4. Model validation

Typical Radiomic workflow
Thawani et al, Lung Cancer - doi.org/10.1016/j.lungcan.2017.10.015

Segmentation

The first step is to segment the lung cancer from the images. We have 957 CT scans across Imperial which have been manually segmented by two expert radiologists, as well as external datasets from e.g. The Cancer Imaging Archive.

For automated segmentation, we trained an nnU-Net model, a deep-learning medical image segmentation tool based on a U-Net architecture
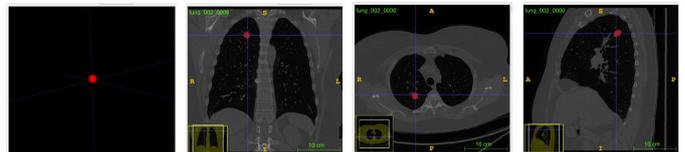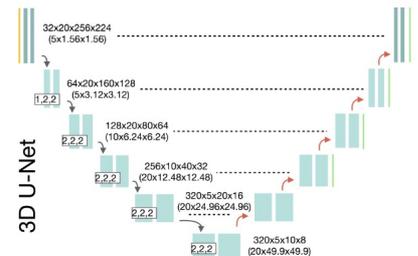

We assessed nnU-Net automated performance stratified against a radiologist "difficulty assessment" for each tumour

Feature Extraction & Selection
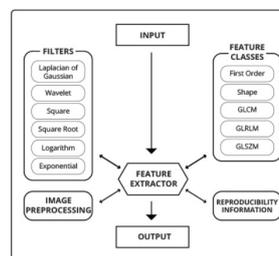
A huge number of quantitative features can then be extracted using software such as PyRadiomics. This captures a range of features, from simple first order statistics and shape based measures to more complex features based on wavelet or Gaussian filters.

These features are statistically compared to actual patient outcomes to select the most important and representative features, and importantly exclude non-reproducible features.
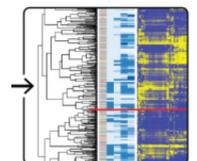
Conclusion and Next Steps

We have successfully designed a fully automated pipeline to carry out the whole workflow for obtaining lung cancer radiomics.

We are continuing to train the model, and validating the radiomic features against local and external data.

As a fully automated pathway, this can now be more readily extended for federated learning across multiple sites and datasets, using platforms such as FLIP (Federated Learning Interoperability Platform) and AIDE (AI Deployment Engine).