Predicting Nodule Malignancy using a CNN Ensemble Approach

Lung cancer is the leading cause of cancer-related deaths globally, which makes early detection and diagnosis a high priority. Computed tomography (CT) is the method of choice for early detection and diagnosis of lung cancer. Radiomics features extracted from CT-detected lung nodules provide a good platform for early detection, diagnosis, and prognosis. In particular when using low dose CT for lung cancer screening, effective use of radiomics can yield a precise non-invasive approach to nodule tracking. Lately, with the advancement of deep learning, convolutional neural networks (CNN) are also being used to analyze lung nodules. In this study, our own trained CNNs, a pre-trained CNN and radiomics features were used for predictive analysis. Using subsets of participants from the National Lung Screening Trial, we investigated if the prediction of nodule malignancy could be further enhanced by an ensemble of classifiers using different feature sets and learning approaches. We extracted probability predictions from our different models on an unseen test set and combined them to generate better predictions. Ensembles were able to yield increased accuracy and area under the receiver operating characteristic curve (AUC). The best-known AUC of 0.96 and accuracy of 89.45% were obtained, which are significant improvements over the previous best AUC of 0.87 and accuracy of 76.79%.