Lung cancer is one of the most common cancers and the first leading cause of people cancer death, which has a serious impact on people’s health. In order to assist doctors improving the accuracy and objectivity of the lung cancer and reducing the misdiagnosis rate of malignant tumors, many computer-aided diagnosis (CAD) methods have been widely used in lung cancer diagnostic process. The Lung Image Database Consortium (LIDC) Data Set is a web-accessible international resource for development, training, and evaluation of CAD methods for lung cancer detection and diagnosis. It consists of diagnostic and lung cancer screening thoracic computed tomography (CT) scans with marked-up annotated lesions.

In this thesis, we propose a novel method to recognize the rate of malignancy of lung cancers utilizing LIDC Data Set. In order to distinguish the ‘semantic gap’ between computer-level image features and human-level semantic features, we provide a quantitative approach for finding the relationships between computer-calculated features and medical semantic concepts used for computer-aided diagnosis characterization system. We chose four semantic descriptors provided by LIDC to predictive values for benign and malignant lesions: sphericity, margin, texture and malignancy. For sake of narrow down the ‘semantic gap’, artificial neural networks (ANNs) is proposed for mapping these two levels of features. Two experiments were performed in this study. Correlation analysis experiment was applied to explore the correlations between the image features and semantic characteristics. Agreement experiment was performed to verify the improvement of quantitative agreement, when we utilized ANNs with composite experts’ opinions and at last two radiologists agreed on a rating for the nodule. Furthermore, in order to reduce the quantity of involved image features, we used the principle component analysis (PCA).