Development of a Self-learning Tool for Radiologists using Content-Based Image Retrieval Techniques

Abstract

Lung cancer accounts for the highest number of cancer-related deaths as compared to other types of cancer in both men and women. Pulmonary nodules are potential manifestations of lung cancer. Pulmonary nodules are blob-like structures with the diameter in the range of 3 mm to 30 mm. Several studies show that screening of lung cancer can substantially reduce the mortality rate. Accurate interpretation of pulmonary nodules is essential for the diagnosis of lung cancer and subsequent plan of treatment. Trainee radiologists have to depend on experienced professionals for interpretation of lung CT images. The lack of time of experienced radiologists is the major bottleneck for such traditional learning procedure. The images and interpretation of similar nodules could assist the trainee radiologists in self-learning as well as in differential diagnosis. Hence, there is a need for the development of a self-learning tool to assist radiologists in learning and differential diagnosis of lung cancer.

The present research work focuses on the development of a self-learning tool using content-based image retrieval techniques. Radiologists could find similar nodules from a database using this tool. Different steps associated with the development of the self-learning tool are the segmentation of pulmonary nodules, representation of nodules using machine level features, computation of similarity, and retrieval of similar nodules. Previously reported retrieval systems of pulmonary nodules can not be put into practice as radiologists need to draw the boundary of nodules during query formation as well as for feature database creation. In the proposed retrieval system, the pulmonary nodules are segmented using a semi-automated technique, which requires a seed point on the nodule from the end-user. The proposed segmentation technique is applicable for all types of pulmonary nodules irrespective of their internal texture (viz. solid, part-solid and non-solid) and external attachment (viz. juxta-pleural and juxta-vascular). Differential-geometry-based techniques are developed for the computation of spiculation, lobulation, and sphericity. Margin sharpness is represented using the combination of the histogram spread of averaged gradient and acutance of the nodule. The performance of the proposed retrieval system is affected by the accuracy of the segmentation technique. Several combinations of shape-based, margin-based, and texture-based features are studied to improve the accuracy of retrieval. Considering city-block distance, the precision achieved by the proposed retrieval system are 85.42%, 82.84%, and 82.29% for 3, 5, and 7 retrieved nodules, respectively. The results of the proposed retrieval system are comparable to those of most recent technique reported in the literature, which depends on radiologists for segmentation of nodules. The proposed tool can be used with minimal user intervention.

A classification scheme is also developed to categorize pulmonary nodules as benign and malignant. The pulmonary nodules are segmented using a semi-automated technique. The efficacies of several combinations of shape-based, margin-based and texture-based features are studied from the classification perspective. A support vector machine is used for differentiation of malignant and benign nodules. Separate training and testing data sets are constructed using a five-fold cross-validation approach. The average value of the area under the receiver operating characteristic curve achieved by the proposed method is 0.9489. Due to the inclusion of more relevant features, the proposed classification scheme outperforms the most recent work reported in the literature, which depends on radiologists for manual segmentation.