

ABSTRACT

Multiclass classification of Non-Small Cell Lung Cancer (NSCLC) based on CT images still faces challenges in terms of feature representation and model generalization. Previous studies have predominantly utilized a single feature extraction method or demonstrated high performance in cross-validation settings but experienced performance degradation when evaluated on independent datasets. This study aims to extract texture feature representations from lung CT images using hybrid feature extraction methods, classify the images into four classes, and identify the optimal combination of hybrid feature extraction techniques and machine learning classification algorithms. This study utilized a dataset of 851 images, which were divided into 70% training data and 30% testing data, with a 5-fold cross-validation technique applied to the training set. The preprocessing stage included image resizing and contrast enhancement using Contrast Limited Adaptive Histogram Equalization (CLAHE). The main focus of the research was the integration of hybrid texture features, namely Gray Level Co-occurrence Matrix (GLCM) + Local Binary Pattern (LBP), Gabor filter + LBP, and Gabor filter + GLCM, which were subsequently classified using K-Nearest Neighbor (KNN) and Random Forest algorithms. The results indicate that the GLCM + LBP feature combination supported by the Random Forest algorithm achieved the best performance with an accuracy of 99.22%. These findings demonstrate that the integration of hybrid texture features and appropriate classification algorithms effectively improves multi-class lung cancer classification accuracy based on CT images.

Keywords: lung cancer, NSCLC, hybrid feature extraction, machine learning