عنوان مقاله:

Statistical Analysis of Haralick Features for Discrimination of Malignancy from Benignity in Breast Ultrasound Images

نویسندگان:

S Aliakbari - *Department of Radiology, Allied Health Sciences Faculty, Semnan University of Medical Sciences, Semnan, Iran,*

P Hejazi - *Department of Medical Physics, Semnan University of Medical Sciences, Semnan, Iran*

N Zolfagharkhani, - *Student search committee Semnan University of Medical Sciences, Semnan, Iran*

خلاصه مقاله:

Introduction: Effective detection of breast cancer is crucial as it is one of the major causes of female mortality. While mammography is the primary screening method, ultrasound is also valuable, especially for young women and those with dense breast tissue [۱]. However, ultrasound imaging has resolution constraints and is susceptible to noise, which makes distinguishing between benign and malignant breast tumors challenging [۲]. Artificial intelligence, specifically deep learning, offers a promising solution through automated image analysis and diagnostic support [۱]. This study aims to evaluate the effectiveness of Haralick texture features that are extracted from ۲D breast ultrasound images by the method called GLCM, in classifying tumors as benign or malignant [۳][۴]. The ultimate goal is to determine the suitability of these features for machine learning applications and comprehensive characterization of breast lesions.Materials and Methods: In this study, breast ultrasound images containing both benign and malignant masses were used as the dataset. The images have been divided into two categories: benign and malignant. Some of the images containing two masses have been removed from the dataset. The remaining images were then examined for size consistency, and those that were not of the same size were excluded from the dataset.The skimage.feature library was used for extracting Haralick features by Python. In this library, a GLCM (Gray-Level Co-occurrence Matrix) matrix was first obtained from the ROI (Region of Interest). Six Haralick features, namely Contrast, Dissimilarity, Homogeneity, ASM (Angular Second Moment), Energy, and Correlation, were extracted at four different angles: ۰, ۴۵, ۹۰, and ۱۳۵ degrees[۶]. Subsequently, the extracted features were averaged for each angle, and these averaged values were used as the final dataset for analysis. Results: The number of benign and malignant data points in this study is ۴۳۷ and ۲۱۱ respectively [۷]. Shapiro-Wilk test was conducted on the Data by Python from breast ultrasound images in two groups, benign and malignant, which showed that the data distribution does not follow a normal distribution [۸]. Therefore, a Mann-Whitney test was used to compare the means of the features [۹]. The Mann-Whitney test was performed for both groups of images, and it was concluded that all six Haralick features were suitable for classification.Conclusion: One of the crucial steps in machine learning is selecting appropriate features for data modelling and classification [۱۰]. One effective approach to finding suitable ...

کلمات کلیدی:
GLCM, Haralick features, Breast Cancer, Python