Abstract:

Cancer remains one of the deadliest diseases worldwide, with breast cancer or ductal carcinoma being a leading cause of death among women. Diagnosis often begins with biopsy tests, among which Fine Needle Aspiration Cytology (FNAC) is a low-cost and rapid method. Cytology image analysis thus plays a vital role in early detection and effective treatment. Traditionally, experts classify samples as benign or malignant by observing cellular characteristics—a process that is time-consuming, expertise-dependent, and prone to human error due to the increasing number of cases.

To address these challenges, this research focuses on developing Computer-Aided Diagnosis (CAD) systems for breast cancer diagnosis. Deep Convolutional Neural Networks (DCNNs) have significantly advanced cytology image analysis, yet existing models face limitations due to insufficient annotated datasets. To overcome this, a novel breast cytology image dataset named JUCYT has been developed, featuring both cancerous and non-cancerous samples from Indian patients.

To enhance dataset diversity, both traditional and deep learning-based augmentation techniques have been proposed, including a fuzzy template-based augmentation method and generative models such as SynCGAN and Br-CytoGAN. While SynCGAN produced low-resolution samples, Br-CytoGAN generated high-quality selective synthetic images, guided by DCNN models like DenseNet-169, InceptionNet-v3, and ResNet-18. Additionally, the study explored unpaired image translation from histopathology to cytology images using CycleGAN and Neural Style Transfer.

For segmentation of cellular objects from breast cytology images, novel deep learning frameworks were introduced. A fuzzy rank-based ensemble combining UNet and SegNet improved accuracy, while a hybrid localization-to-segmentation model using Faster-RCNN and MedSAM further enhanced segmentation performance.

For classification, both traditional machine learning and deep learning approaches were evaluated. Cellular features were used with KNN, Decision Tree, and SVM classifiers, while fine-tuned CNNs (ResNet, DenseNet, InceptionNet) achieved superior performance. A Gaussian copula-based CNN ensemble (GC-EnC) and an Artificial Electric Field algorithm for optimal feature selection further boosted accuracy. Synthetic samples generated by GAN models were also used for data augmentation.

Finally, an automated web application was developed for real-time breast cytology diagnosis, offering pathologists faster and more reliable decision support. Overall, this research demonstrates the efficacy of deep learning in improving breast cancer diagnostics and establishes a strong foundation for future cytology image analysis advancements.