**BTS-ST: Swin transformer network for segmentation and classification of multimodality breast cancer images**

Lung cancer is the cancer that is diagnosed the second most frequently worldwide, after breast cancer. Breast cancer analysis employing ultrasound, mammography, and MRI modalities as the initial screening method for the early diagnosis of breast tumours in women. The correct segmentation of breast tumours is still a challenging task because of the random diversity, atypical forms, and hazy boundaries of tumour zones. The existing convolutional neural networks (CNNs) inherit their limitations by extracting global context information and, in most cases, were ineffective at producing results that were adequate.

We came up with the BTS-ST network, a brand-new approach for classifying and segmenting breast tumours that is motivated by Swin-Transformer (ST). In order to enhance the capabilities of its typical CNNs-based U-Net, the BTS-ST network adds Swin-Transformer. To enhanceIn order to enhance the capabilities of its typical CNNs-based U-Net, the BTS-ST network adds Swin-Transformer. We initially created a Spatial Interaction block (SIB), encoding spatial knowledge in the Swin Transformer block by creating pixel-level correlation, to increase the feature representation capabilities of tumours with irregular shapes.

Building a Feature Compression block (FCB) to prevent information loss and compress smaller-scale features in patch token down sampling of Swin-Transformer improves the segmentation accuracy of small-scale tumour regions. In order to incorporate global dependencies from Swin-Transformer into the features from CNN hierarchically, a Relationship Aggregation block (RAB) is created as a bridge between dual encoders. Using multimodality ultrasound, mammogram, and MRI-based datasets, extensive studies are conducted on the segmentation and classification of breast tumours. The outcomes show that our suggested.

In this research, we present a network based on the BTS-ST vision transformer for segmenting and categorising breast tumour images. Spatial Interaction Block (SIB), Feature Compression Block (FCB), and Relationship Aggregation Block (RAB) are the three main building blocks that make up the core framework of BTS-ST. We have carried out numerous experiments on MRI, ultrasound, and mammogram pictures to demonstrate the model's efficacy on multimodal imaging. The proposed network performed well on the Private BUS dataset, RIDER Breast dataset, and CBIS-DDSM Mammogram dataset, with F1 values of (0.908), (0.833), and (0.771), respectively. In order to accurately anticipate breast tumours in women before they manifest, we think that our suggested network can be employed as a computer-aided breast tumour screening method.

The attention mechanism often occupies most of the processing time of the visual Transformer.