**SAMASK-CLTR : A spatial-aware mask guided**

**learning model for benign and malignant tumor**

**classification in ABUS/ABVS**

PEIRONG XU

we propose SAMASK-CLTR (Spatial-Aware Mask Prompting with Con-

volutional Transformer Architecture), a hybrid framework that combines the feature

extraction power of CNNs with the global modeling capability of Transformers. In

our approach, ResNet-50 extracts hierarchical, multi-scale features that are refined

by a Transformer encoder-decoder to capture global context. Crucially, during de-

coding, a mask prompt enhanced with 3D positional encoding guides the network to

focus on key tumor regions, directly addressing the challenges of precise localiza-

tion and classification. Experiments on 6,973 ABUS/ABVS images—including 6,873

clinical cases from Internal Datasets and 100 cases from the public ABUS Challenge

Cup—demonstrate that SAMASK-CLTR achieves AUCs of 88.45% and 70.46% on

internal and external datasets, respectively. These results highlight the potential of our

framework to significantly enhance breast cancer diagnosis by improving the accuracy and reliability of lesion classification