**TransMed: Transformers Advance Multi-modal Medical Image Classification (Paper Review)**

* Over the past decade, convolutional neural networks (CNN) have shown very competitive performance in medical image analysis tasks, such as disease classification, tumour segmentation, and lesion detection. CNN has great advantages in extracting local features of images.
* However, due to the locality of convolution operation, it cannot deal with long-range relationships well.
* Recently, transformers have been applied to computer vision and achieved remarkable success in large-scale datasets. Compared with natural images, multi-modal medical images have explicit and important long-range dependencies, and effective multi-modal fusion strategies can greatly improve the performance of deep models.
* This prompts us to study transformer-based structures and apply them to multi-modal medical images. Existing transformer-based network architectures require large-scale datasets to achieve better performance.
* However, medical imaging datasets are relatively small, which makes it difficult to apply pure transformers to medical image analysis.
* Therefore, we propose TransMed for multi-modal medical image classification. TransMed combines the advantages of CNN and transformer to efficiently extract low-level features of images and establish long-range dependencies between modalities.
* We evaluated our model for the challenging problem of preoperative diagnosis of parotid gland tumours, and the experimental results show the advantages of our proposed method. We argue that the combination of CNN and transformer has tremendous potential in a large number of medical image analysis tasks. To our best knowledge, this is the first work to apply transformers to medical image classification.