The paper we chose to present is titled \*\*"Deep Evidential Fusion with Uncertainty Quantification for Multimodal Medical Image Segmentation."\*\*

In clinical practice, various imaging modalities, such as PET, CT, or MRI sequences, are often combined to provide a more comprehensive understanding of a patient's condition. For instance, PET images are excellent for highlighting metabolic activity, while CT images reveal anatomical structures more clearly. By combining these different modalities, we can improve diagnostic accuracy and achieve a deeper understanding of the condition.

However, there's a significant challenge with this approach. Traditional segmentation methods for multimodal images often assume that all data sources are equally reliable. But in practice, each modality may vary in terms of quality, resolution, and reliability. If we assume that all modalities are equally trustworthy, it can lead to segmentation errors, ultimately impacting the diagnosis and treatment.

This paper aims to address these issues by introducing a \*\*deep evidential fusion framework\*\*. The proposed framework leverages \*\*Dempster-Shafer Theory\*\* to quantify both the uncertainty and the reliability of each modality, which aims for more accurate and explainable segmentation results. This approach helps us to make better, more informed clinical decisions.

## Proposed Framework (5 minutes)

The proposed framework can be broken down into three main modules:

- \*\*Feature Extraction (FE)\*\*

- \*\*Evidence Mapping (EM)\*\*

- \*\*Multi-modality Evidence Fusion (MMEF)\*\*

And Each imaging modality has its own FE and EM modules.

Now Let's look into each of these components in more detail.

## Feature Extraction (FE)

The first module is the \*\*Feature Extraction\*\* (FE) module. The framework uses well-known deep learning models like \*\*UNet, nnUNet\*\*, or \*\*nnFormer\*\* to extract relevant features from each modality independently.

For example, if we input a 128x128x128 single-channel grayscale image into the FE module, it will output a feature representation of the same spatial size, but with \*\*H channels\*\* instead of a single channel. This means we are now working with \*\*H-dimensional features\*\*, where H represents the number of features extracted at each voxel.

In this paper, \*\*H is set to 2\*\* for the PET-CT lymphoma dataset and \*\* 4\*\* for the multi-MRI BraTS2021 dataset.

## Evidence Mapping (EM)

After feature extraction, the extracted features are transformed into \*\*mass functions\*\* using the \*\*Evidential Neural Network (ENN)\*\* as the Evidence Mapping (EM) module. The mass functions represent the evidence for each voxel, resulting in a tensor of size \*\*128x128x128x(K+1)\*\*, where \*\*K+1\*\* represents the number of classes plus one for the \*\*frame of discernment\*\*.

The \*\*ENN module\*\* is structured with three main layers: a \*\*prototype activation layer\*\*, a \*\*mass calculation layer\*\*, and a \*\*combination layer\*\*.

- \*\*Prototypes\*\*: Prototypes in the ENN are obtained by running the k-means algorithm on the features extracted by the FE module. For the PET-CT lymphoma dataset, we use \*\*10 prototypes\*\*, while for the multi-MRI BraTS2021 dataset, we use \*\*20 prototypes\*\*.

- \*\*Similarity Calculation\*\*: In the first hidden layer, the similarity s\_i^t between an input feature vector x and a prototype $p\_i^t$ is calculated using an activation function.

- The activation of unit \*\*i\*\* in the prototype layer is defined as follows:

$$

- Where, \*\*γ\_i^t\*\* and \*\*α\_i^t\*\* are learnable parameters.

- \*\*Mass Function Calculation\*\*: Next, a second hidden layer computes the \*\*mass function\*\* \*\*\(m\_i^t\)\*\*, which represents the belief provided by prototype \*\*\(p\_i^t\)\*\*. The focal sets of \*\*\(m\_i^t\)\*\* are singletons \*\*\(θ\_k\)\*\* (\(k = 1, \dots, K\)) and the universal set \*\*Θ\*\*. The mass function is defined as follows:

$$

- where \*\*\(u\_{ik}^t\)\*\* is a learnable parameter that represents the membership degree of prototype \*\*\(p\_i^t\)\*\* to class \*\*\(θ\_k\)\*\*. In this paper, it is initialized by drawing from uniform random numbers and normalizing them.

- \*\*Fusing Mass Functions\*\*: Finally, in the third hidden layer, the \*\*I\*\* mass functions are combined using \*\*Dempster's rule\*\* to provide a comprehensive opinion on the fused mass function \*\*\(m\)\*\* for the input feature vector \*\*\(x\)\*\*.

## Multi-modality Evidence Fusion (MMEF)

Once we have evidence from each modality, the next step is to \*\*fuse\*\* them. Instead of fusing at the mass function level, the \*\*MMEF module\*\* fuses them at the \*\*contour function level\*\*. This approach decreases the computing costs and facilitates \*\*plausibility-probability transformation\*\* and makes it easier to quantify the reliability of each modality.

The reliability of each modality is represented by a vector \*\*β\*\*, which indicates how reliable a particular modality is for a given class. In this paper, the reliability coefficients are initialized to \*\*0.5\*\*.

The following are the mathematical expressions for the MMEF process:...

## Loss Function (30 seconds)

When it comes to the optimisation of the framework, the authors opted to minimise a combined loss function defined as follows:

$$

loss = loss\_s + loss\_f

$$

- \*\*Loss\_s\*\* evaluates the segmentation performance of each modality individually and aggregates the results.

- \*\*Loss\_f\*\* evaluates the segmentation performance after combining all modalities, thus ensuring that the fusion process leads to an overall improvement.

Note that \*\*\(N\)\*\* is the number of voxels, and \*\*\(G\_{kn}\)\*\* represents whether voxel \*\*n\*\* belongs to class \*\*\(θ\_k\)\*\*.

## Training Process

To recap,

The learnable parameters in this framework are:

- \*\*Feature Extraction (FE) module\*\*: Weights of the deep learning models.

- \*\*Evidence Mapping (EM) module\*\*: Parameters \*\*\(α\_i^t, γ\_i^t, u\_{ik}^t\)\*\*.

- \*\*Multi-modality Evidence Fusion (MMEF) module\*\*: Reliability coefficients \*\*\(β\)\*\*.

The training process is carried out in three main steps:

1. first，Each modality's FE module is trained independently to extract features.

2. then，After training the FE modules, their weights are fixed, and the EM and MMEF modules are optimised.

3. Finally, the entire framework, including the FE, EM, and MMEF modules, is fine-tuned for a few epochs.