JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

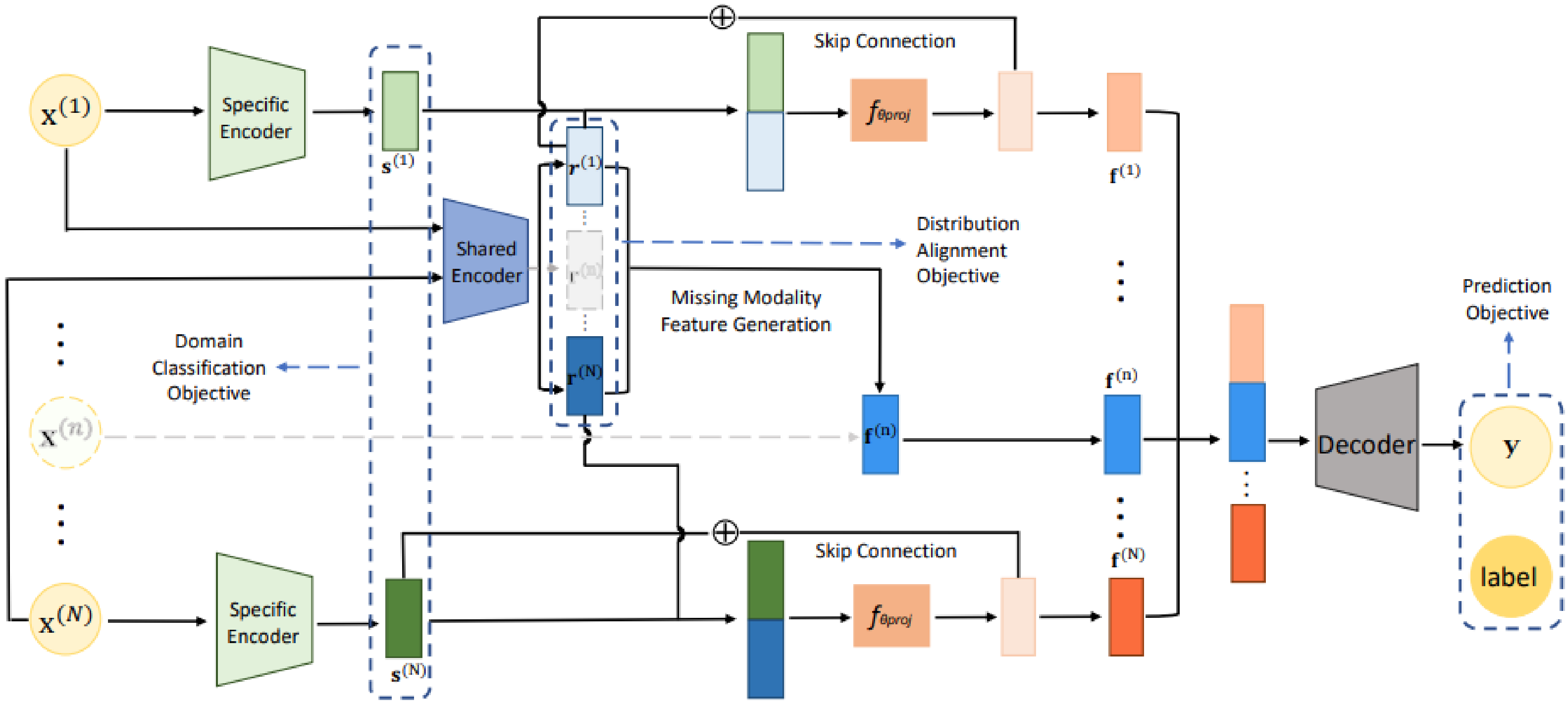
Multi-modal Learning with Missing Modality via Shared-Specific Feature Modelling

Hu Wang , Yuanhong Chen , Congbo Ma , Jodie Avery,
Louise Hull and Gustavo Carneiro

The University of Adelaide

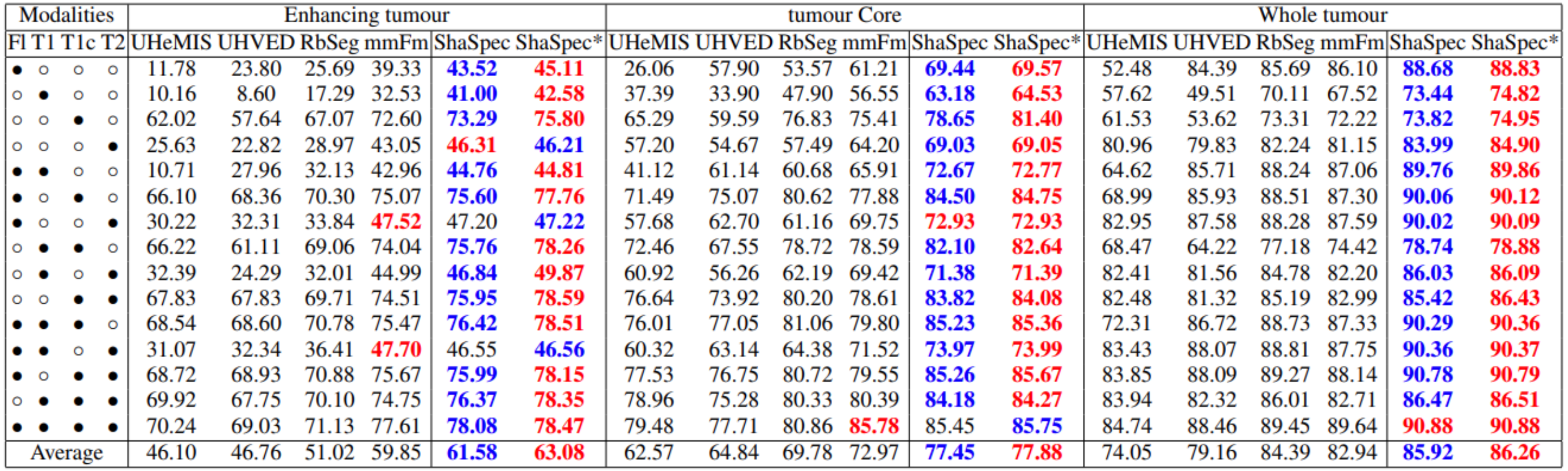
Paper Tag: WED-PM-336

A Quick-look of the Paper



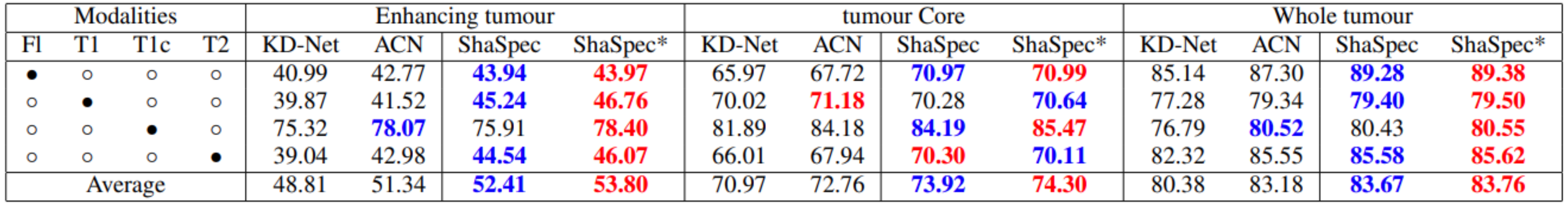
We propose the Shared-Specific Feature Modelling (ShaSpec) method that is considerably simpler and more effective than competing approaches that address the issues above. Also, the design simplicity of ShaSpec enables its easy adaptation to multiple tasks, such as classification and segmentation.

Results --- A Quick Look



Modalities				Enhancing tumour						tumour Core						Whole tumour					
Fl	T1	T1c	T2	UHeMIS	UHVED	RbSeg	mmFm	ShaSpec	ShaSpec*	UHeMIS	UHVED	RbSeg	mmFm	ShaSpec	ShaSpec*	UHeMIS	UHVED	RbSeg	mmFm	ShaSpec	ShaSpec*
●	○	○	○	11.78	23.80	25.69	39.33	**43.52**	**45.11**	26.06	57.90	53.57	61.21	**69.44**	**69.57**	52.48	84.39	85.69	86.10	**88.68**	**88.83**
○	●	○	○	10.16	8.60	17.29	32.53	**41.00**	**42.58**	37.39	33.90	47.90	56.55	**63.18**	**64.53**	57.62	49.51	70.11	67.52	**73.44**	**74.82**
○	○	●	○	62.02	57.64	67.07	72.60	**73.29**	**75.80**	65.29	59.59	76.83	75.41	**78.65**	**81.40**	61.53	53.62	73.31	72.22	**73.82**	**74.95**
○	○	○	●	25.63	22.82	28.97	43.05	**46.31**	**46.21**	57.20	54.67	57.49	64.20	**69.03**	**69.05**	80.96	79.83	82.24	81.15	**83.99**	**84.90**
●	●	○	○	10.71	27.96	32.13	42.96	**44.76**	**44.81**	41.12	61.14	60.68	65.91	**72.67**	**72.77**	64.62	85.71	88.24	87.06	**89.76**	**89.86**
●	○	●	○	66.10	68.36	70.30	75.07	**75.60**	**77.76**	71.49	75.07	80.62	77.88	**84.50**	**84.75**	68.99	85.93	88.51	87.30	**90.06**	**90.12**
●	○	○	●	30.22	32.31	33.84	**47.52**	47.20	**47.22**	57.68	62.70	61.16	69.75	**72.93**	**72.93**	82.95	87.58	88.28	87.59	**90.02**	**90.09**
○	●	●	○	66.22	61.11	69.06	74.04	**75.76**	**78.26**	72.46	67.55	78.72	78.59	**82.10**	**82.64**	68.47	64.22	77.18	74.42	**78.74**	**78.88**
○	●	○	●	32.39	24.29	32.01	44.99	**46.84**	**49.87**	60.92	56.26	62.19	69.42	**71.38**	**71.39**	82.41	81.56	84.78	82.20	**86.03**	**86.09**
○	○	●	●	67.83	67.83	69.71	74.51	**75.95**	**78.59**	76.64	73.92	80.20	78.61	**83.82**	**84.08**	82.48	81.32	85.19	82.99	**85.42**	**86.43**
●	●	●	○	68.54	68.60	70.78	75.47	**76.42**	**78.51**	76.01	77.05	81.06	79.80	**85.23**	**85.36**	72.31	86.72	88.73	87.33	**90.29**	**90.36**
●	●	○	●	31.07	32.34	36.41	**47.70**	46.55	**46.56**	60.32	63.14	64.38	71.52	**73.97**	**73.99**	83.43	88.07	88.81	87.75	**90.36**	**90.37**
●	○	●	●	68.72	68.93	70.88	75.67	**75.99**	**78.15**	77.53	76.75	80.72	79.55	**85.26**	**85.67**	83.85	88.09	89.27	88.14	**90.78**	**90.79**
○	●	●	●	69.92	67.75	70.10	74.75	**76.37**	**78.35**	78.96	75.28	80.33	80.39	**84.18**	**84.27**	83.94	82.32	86.01	82.71	**86.47**	**86.51**
●	●	●	●	70.24	69.03	71.13	77.61	**78.08**	**78.47**	79.48	77.71	80.86	**85.78**	85.45	**85.75**	84.74	88.46	89.45	89.64	**90.88**	**90.88**
Average				46.10	46.76	51.02	59.85	**61.58**	**63.08**	62.57	64.84	69.78	72.97	**77.45**	**77.88**	74.05	79.16	84.39	82.94	**85.92**	**86.26**

Table 1. Model performance comparison of **segmentation** Dice score (normalised to 100%) on BraTS2018 of **non-dedicated training**. ShaSpec and ShaSpec* are the proposed models, with ShaSpec* being the model with prediction smoothness enhancement. The best and second best results for each column within a certain type of tumour are in **red** and **blue**, respectively.



Modalities				Enhancing tumour				tumour Core				Whole tumour			
Fl	T1	T1c	T2	KD-Net	ACN	ShaSpec	ShaSpec*	KD-Net	ACN	ShaSpec	ShaSpec*	KD-Net	ACN	ShaSpec	ShaSpec*
●	○	○	○	40.99	42.77	**43.94**	**43.97**	65.97	67.72	**70.97**	**70.99**	85.14	87.30	**89.28**	**89.38**
○	●	○	○	39.87	41.52	**45.24**	**46.76**	70.02	**71.18**	70.28	**70.64**	77.28	79.34	**79.40**	**79.50**
○	○	●	○	75.32	**78.07**	75.91	**78.40**	81.89	84.18	**84.19**	**85.47**	76.79	**80.52**	80.43	**80.55**
○	○	○	●	39.04	42.98	**44.54**	**46.07**	66.01	67.94	**70.30**	**70.11**	82.32	85.55	**85.58**	**85.62**
Average				48.81	51.34	**52.41**	**53.80**	70.97	72.76	**73.92**	**74.30**	80.38	83.18	**83.67**	**83.76**

Table 2. Model performance comparison of **segmentation** Dice score (normalised to 100%) on BraTS2018 of **dedicated training**.

Experiments are conducted on both medical image segmentation and computer vision classification, with results indicating that ShaSpec outperforms competing methods by a large margin. For instance, on BraTS2018, ShaSpec improves the SOTA by more than 3% for enhancing tumour, 5% for tumour core and 3% for whole tumour.

Main Contributions

- An extremely simple yet effective multi-modal learning with missing modality method, called Shared-Specific Feature Modelling (ShaSpec), which is based on modelling and fusing shared and specific features to deal with missing modality in training and evaluation and with dedicated and non-dedicated training;
- To the best of our knowledge, the proposed ShaSpec is the first missing modality multi-modal approach that can be easily adapted to both classification and segmentation tasks given the simplicity of its design.
- Our results on computer vision classification and medical imaging segmentation benchmarks show that ShaSpec achieves state-of-the-art performance. Notably, compared with recently proposed competing approaches on BraTS2018, our model shows segmentation accuracy improvements of more than 3% for enhancing tumour, 5% for tumour core and 3% for whole tumour.

Structure of ShaSpec model with Full modalities

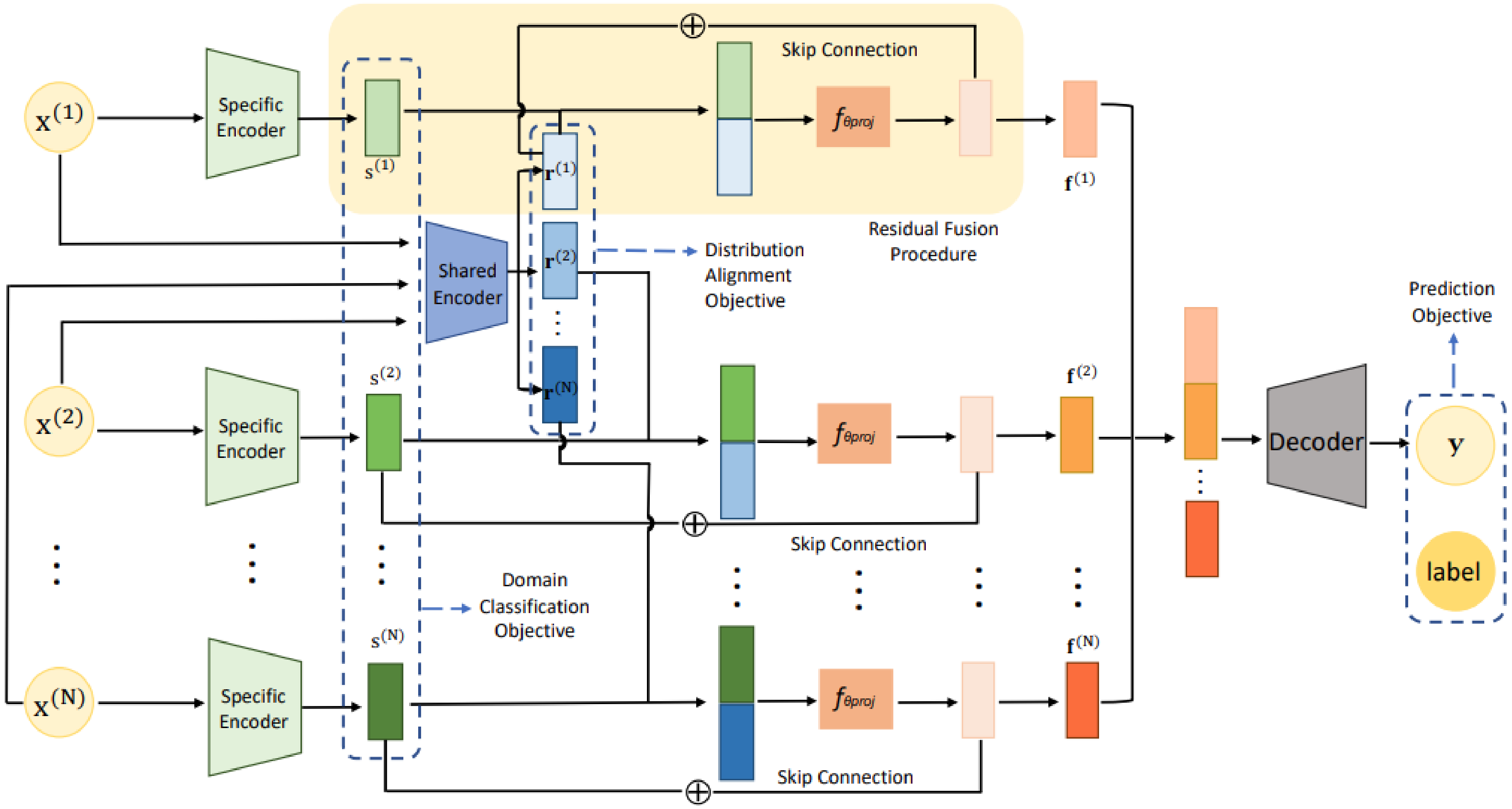


Figure 1. **Full-modality training and evaluation of ShaSpec.** All modalities $\{\mathbf{x}^{(i)}\}_{i=1}^{N} \in \mathcal{M}$ are passed through one shared encoder and individual specific encoders to produce the shared features $\{\mathbf{r}^{(i)}\}_{i=1}^{N}$ and specific features $\{\mathbf{s}^{(i)}\}_{i=1}^{N}$, respectively. Then, in a residual learning manner, the shared and specific features are fused with a linear projection $f_{\theta_{\text{proj}}}(\cdot)$ to get the fused features $\{\mathbf{f}^{(i)}\}_{i=1}^{N}$ for decoding. The dashed blue arrows indicate different objective functions.

Structure of ShaSpec model with missing modalities

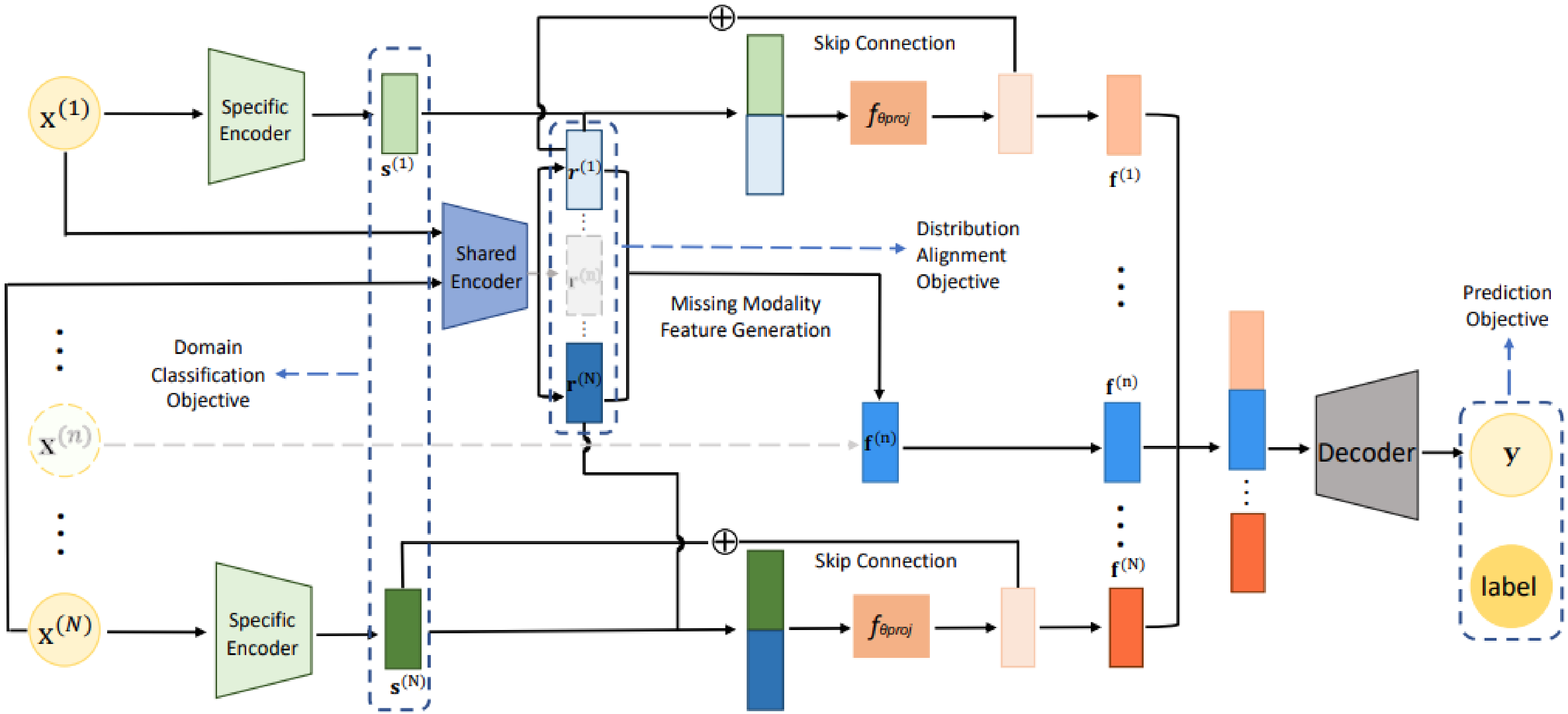
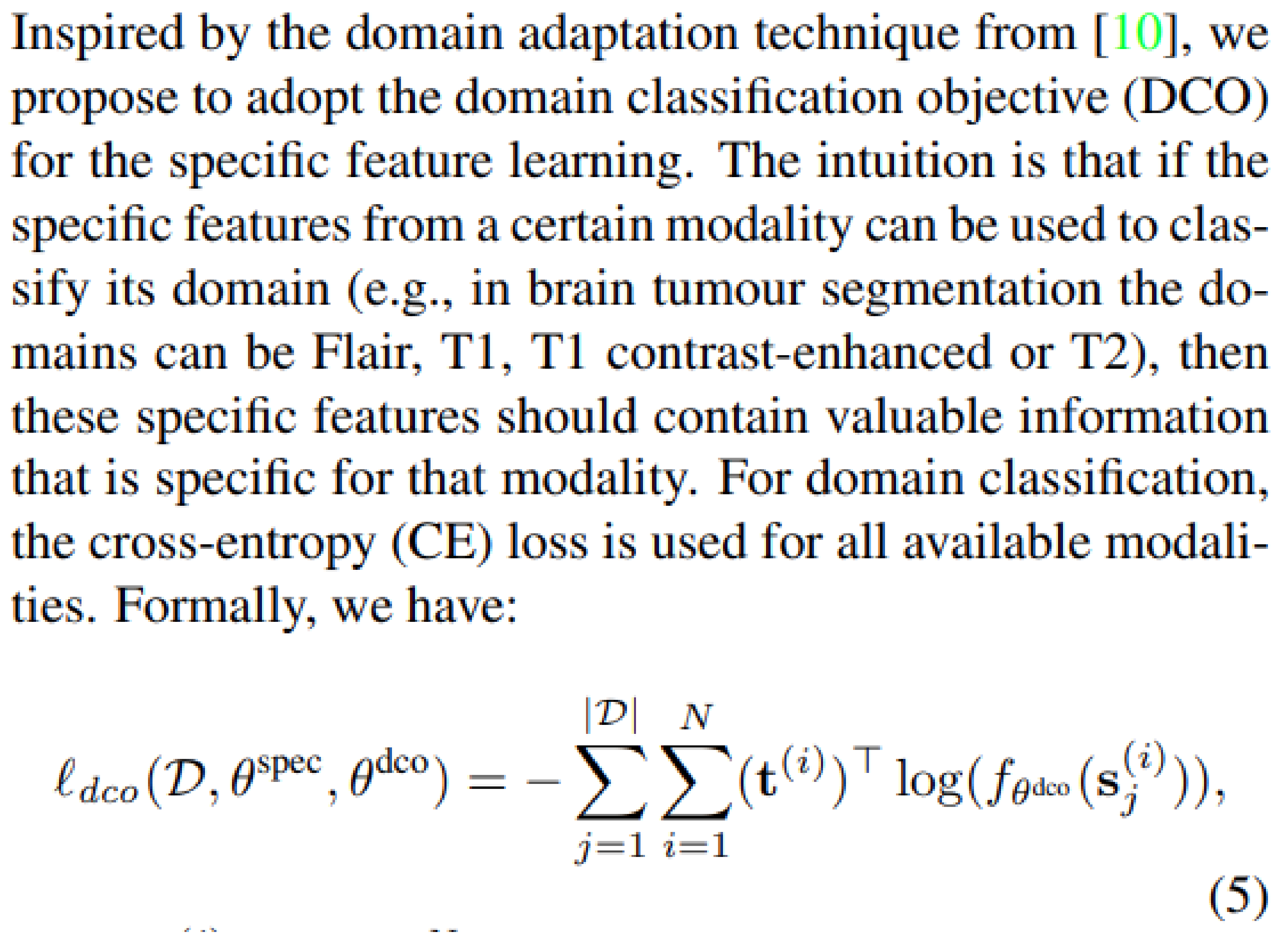


Figure 2. **Missing-modality training and evaluation of ShaSpec.** Without losing generality, we assume $\mathbf{x}^{(n)}$ is missing, where n can be $1, 2, ..., N$. For available modalities $\mathbf{x}^{(1)}, ..., \mathbf{x}^{(n-1)}, \mathbf{x}^{(n+1)}, ..., \mathbf{x}^{(N)}$, the shared-specific fused features $\mathbf{f}^{(1)}, ..., \mathbf{f}^{(n-1)}, \mathbf{f}^{(n+1)}, ..., \mathbf{f}^{(N)}$ are extracted in the same way as in full modality. But for the missing modality data $\mathbf{x}^{(n)}$, the fused features $\mathbf{f}^{(n)}$ are generated from available shared features $\mathbf{r}^{(1)}, ..., \mathbf{r}^{(n-1)}, \mathbf{r}^{(n+1)}, ..., \mathbf{r}^{(N)}$ via a missing modality feature generation process. The dashed blue arrows indicate different objective functions.

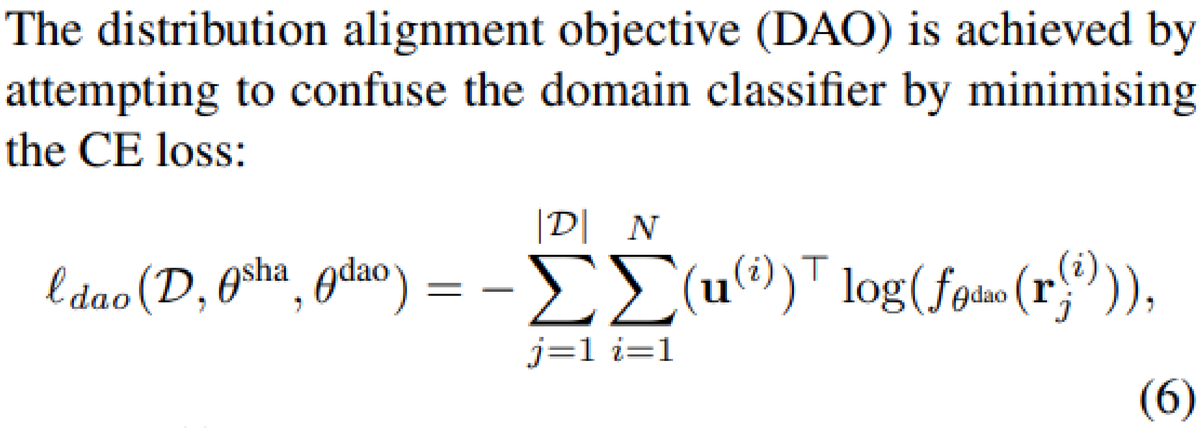
Domain Classification Objective --- Specific Feature Learning



Inspired by the domain adaptation technique from [10], we propose to adopt the domain classification objective (DCO) for the specific feature learning. The intuition is that if the specific features from a certain modality can be used to classify its domain (e.g., in brain tumour segmentation the domains can be Flair, T1, T1 contrast-enhanced or T2), then these specific features should contain valuable information that is specific for that modality. For domain classification, the cross-entropy (CE) loss is used for all available modalities. Formally, we have:

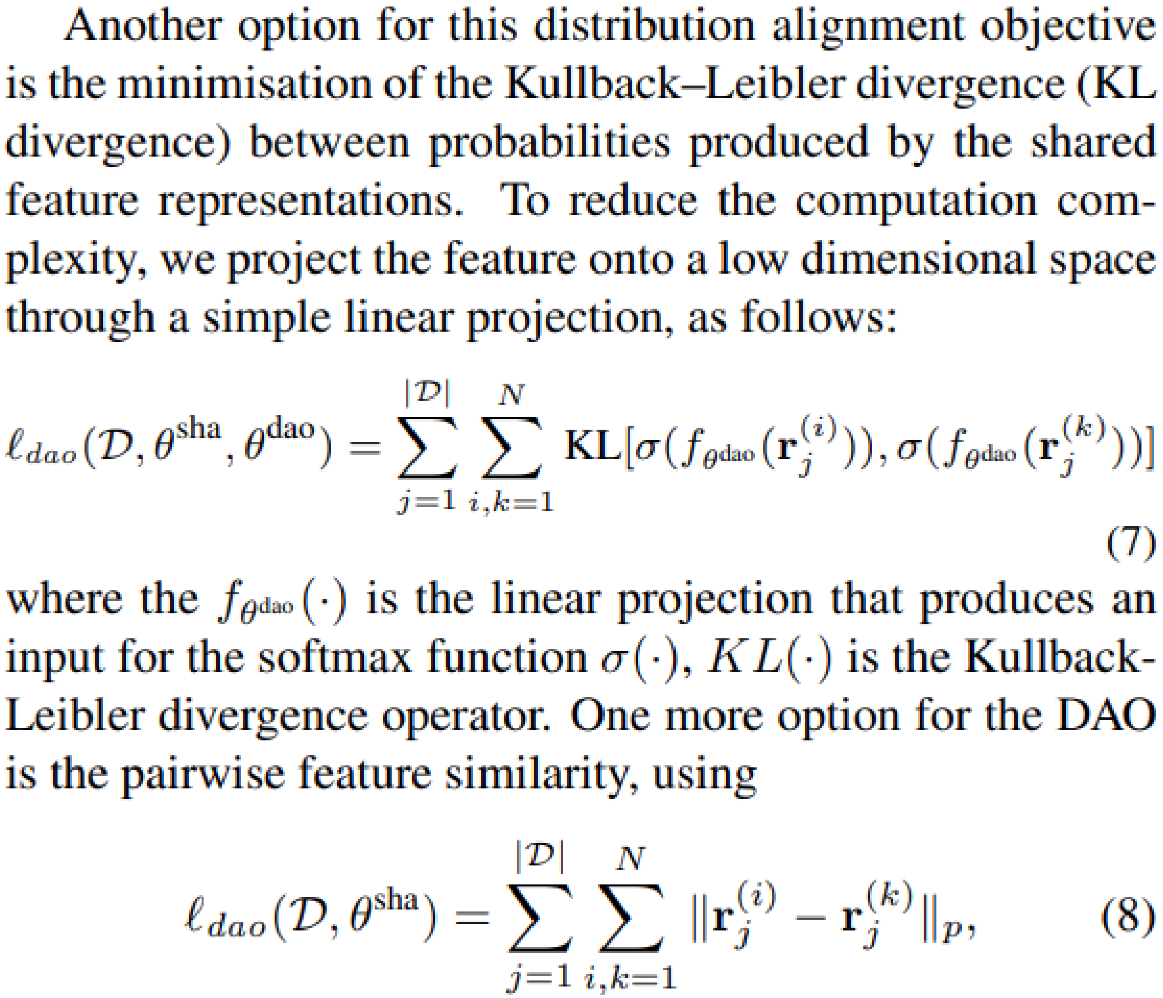
$$\ell_{dco}(\mathcal{D},\theta^{\text{spec}},\theta^{\text{dco}}) = -\sum_{j=1}^{|\mathcal{D}|}\sum_{i=1}^{N}(\mathbf{t}^{(i)})^{\top}\log(f_{\theta^{\text{dco}}}(\mathbf{s}_j^{(i)})), \tag{5}$$

Distribution Alignment Objective --- Shared Feature Learning



The distribution alignment objective (DAO) is achieved by attempting to confuse the domain classifier by minimising the CE loss:

$$\ell_{dao}(\mathcal{D},\theta^{\text{sha}},\theta^{\text{dao}}) = -\sum_{j=1}^{|\mathcal{D}|}\sum_{i=1}^{N}(\mathbf{u}^{(i)})^{\top}\log(f_{\theta^{\text{dao}}}(\mathbf{r}_j^{(i)})), \tag{6}$$



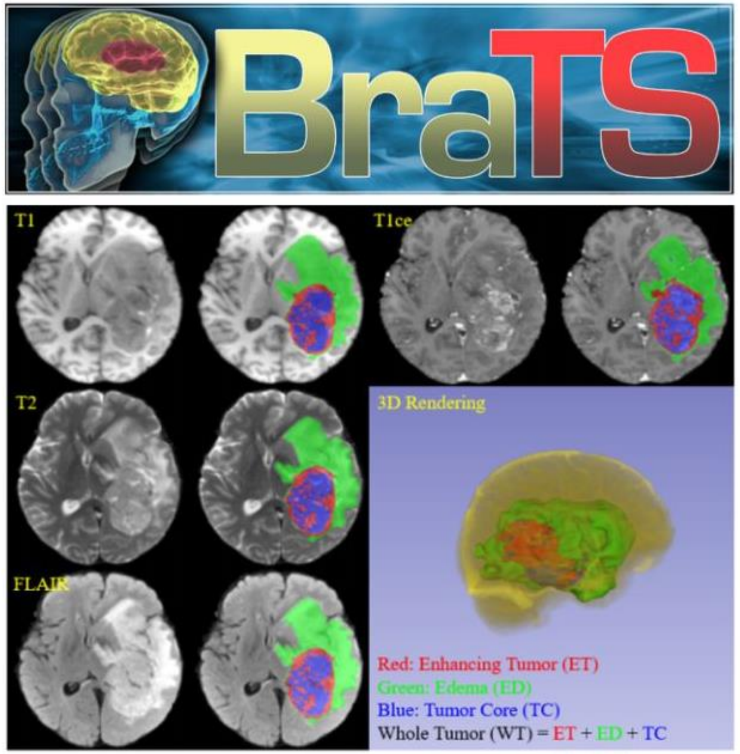
Another option for this distribution alignment objective is the minimisation of the Kullback–Leibler divergence (KL divergence) between probabilities produced by the shared feature representations. To reduce the computation complexity, we project the feature onto a low dimensional space through a simple linear projection, as follows:

$$\ell_{dao}(\mathcal{D},\theta^{\text{sha}},\theta^{\text{dao}}) = \sum_{j=1}^{|\mathcal{D}|}\sum_{i,k=1}^{N}\text{KL}[\sigma(f_{\theta^{\text{dao}}}(\mathbf{r}_j^{(i)})),\sigma(f_{\theta^{\text{dao}}}(\mathbf{r}_j^{(k)}))] \tag{7}$$

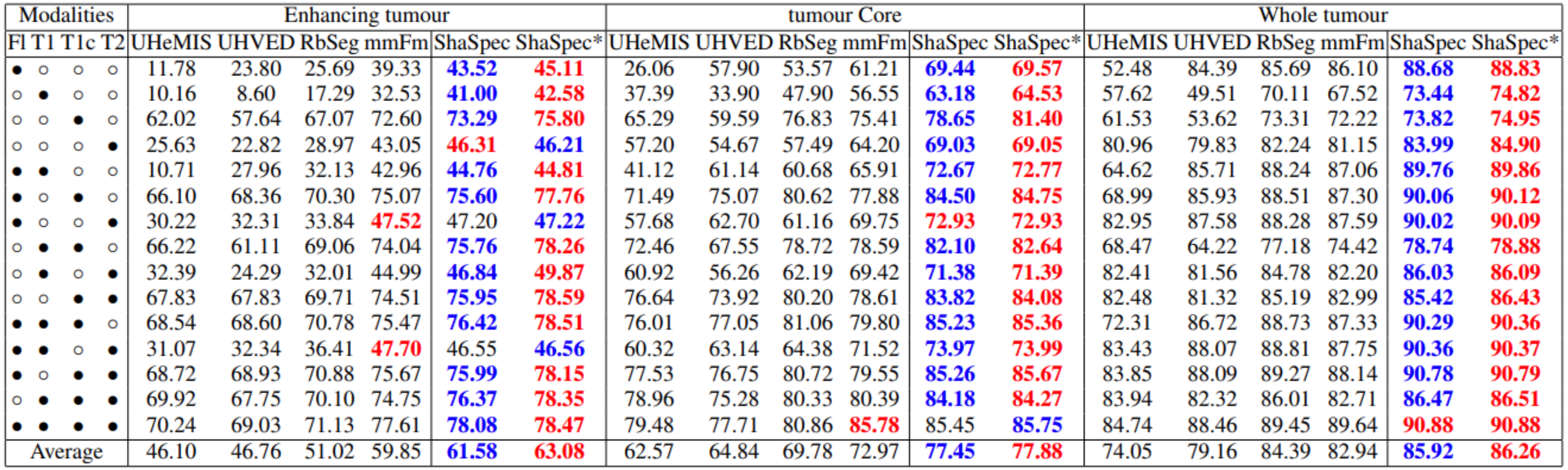
where the $f_{\theta^{\text{dao}}}(\cdot)$ is the linear projection that produces an input for the softmax function $\sigma(\cdot)$, $KL(\cdot)$ is the Kullback-Leibler divergence operator. One more option for the DAO is the pairwise feature similarity, using

$$\ell_{dao}(\mathcal{D},\theta^{\text{sha}}) = \sum_{j=1}^{|\mathcal{D}|}\sum_{i,k=1}^{N}\|\mathbf{r}_j^{(i)} - \mathbf{r}_j^{(k)}\|_p, \tag{8}$$

Dataset & Results --- BraTS2018

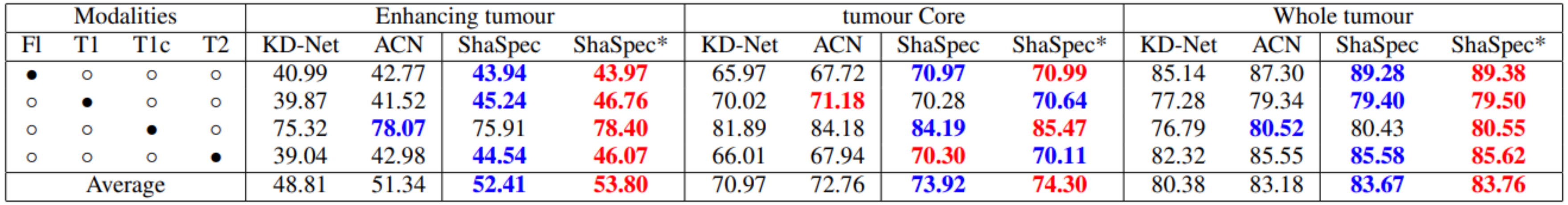


Annotations comprise 3 types of labels: the GD-enhancing tumor (**ET — label 4**), the peritumoral edema (**ED — label 2**), and the necrotic and non-enhancing tumor core (**NCR/NET — label 1**). **285 cases** for training + validation; **66 cases** for testing



Modalities				Enhancing tumour						tumour Core						Whole tumour					
Fl	T1	T1c	T2	UHeMIS	UHVED	RbSeg	mmFm	ShaSpec	ShaSpec*	UHeMIS	UHVED	RbSeg	mmFm	ShaSpec	ShaSpec*	UHeMIS	UHVED	RbSeg	mmFm	ShaSpec	ShaSpec*
●	○	○	○	11.78	23.80	25.69	39.33	**43.52**	**45.11**	26.06	57.90	53.57	61.21	**69.44**	**69.57**	52.48	84.39	85.69	86.10	**88.68**	**88.83**
○	●	○	○	10.16	8.60	17.29	32.53	**41.00**	**42.58**	37.39	33.90	47.90	56.55	**63.18**	**64.53**	57.62	49.51	70.11	67.52	**73.44**	**74.82**
○	○	●	○	62.02	57.64	67.07	72.60	**73.29**	**75.80**	65.29	59.59	76.83	75.41	**78.65**	**81.40**	61.53	53.62	73.31	72.22	**73.82**	**74.95**
○	○	○	●	25.63	22.82	28.97	43.05	**46.31**	**46.21**	57.20	54.67	57.49	64.20	**69.03**	**69.05**	80.96	79.83	82.24	81.15	**83.99**	**84.90**
●	●	○	○	10.71	27.96	32.13	42.96	**44.76**	**44.81**	41.12	61.14	60.68	65.91	**72.67**	**72.77**	64.62	85.71	88.24	87.06	**89.76**	**89.86**
●	○	●	○	66.10	68.36	70.30	75.07	**75.60**	**77.76**	71.49	75.07	80.62	77.88	**84.50**	**84.75**	68.99	85.93	88.51	87.30	**90.06**	**90.12**
●	○	○	●	30.22	32.31	33.84	**47.52**	47.20	**47.22**	57.68	62.70	61.16	69.75	**72.93**	**72.93**	82.95	87.58	88.28	87.59	**90.02**	**90.09**
○	●	●	○	66.22	61.11	69.06	74.04	**75.76**	**78.26**	72.46	67.55	78.72	78.59	**82.10**	**82.64**	68.47	64.22	77.18	74.42	**78.74**	**78.88**
○	●	○	●	32.39	24.29	32.01	44.99	**46.84**	**49.87**	60.92	56.26	62.19	69.42	**71.38**	**71.39**	82.41	81.56	84.78	82.20	**86.03**	**86.09**
○	○	●	●	67.83	67.83	69.71	74.51	**75.95**	**78.59**	76.64	73.92	80.20	78.61	**83.82**	**84.08**	82.48	81.32	85.19	82.99	**85.42**	**86.43**
●	●	●	○	68.54	68.60	70.78	75.47	**76.42**	**78.51**	76.01	77.05	81.06	79.80	**85.23**	**85.36**	72.31	86.72	88.73	87.33	**90.29**	**90.36**
●	●	○	●	31.07	32.34	36.41	**47.70**	46.55	**46.56**	60.32	63.14	64.38	71.52	**73.97**	**73.99**	83.43	88.07	88.81	87.75	**90.36**	**90.37**
●	○	●	●	68.72	68.93	70.88	75.67	**75.99**	**78.15**	77.53	76.75	80.72	79.55	**85.26**	**85.67**	83.85	88.09	89.27	88.14	**90.78**	**90.79**
○	●	●	●	69.92	67.75	70.10	74.75	**76.37**	**78.35**	78.96	75.28	80.33	80.39	**84.18**	**84.27**	83.94	82.32	86.01	82.71	**86.47**	**86.51**
●	●	●	●	70.24	69.03	71.13	77.61	**78.08**	**78.47**	79.48	77.71	80.86	**85.78**	85.45	**85.75**	84.74	88.46	89.45	89.64	**90.88**	**90.88**
Average				46.10	46.76	51.02	59.85	**61.58**	**63.08**	62.57	64.84	69.78	72.97	**77.45**	**77.88**	74.05	79.16	84.39	82.94	**85.92**	**86.26**

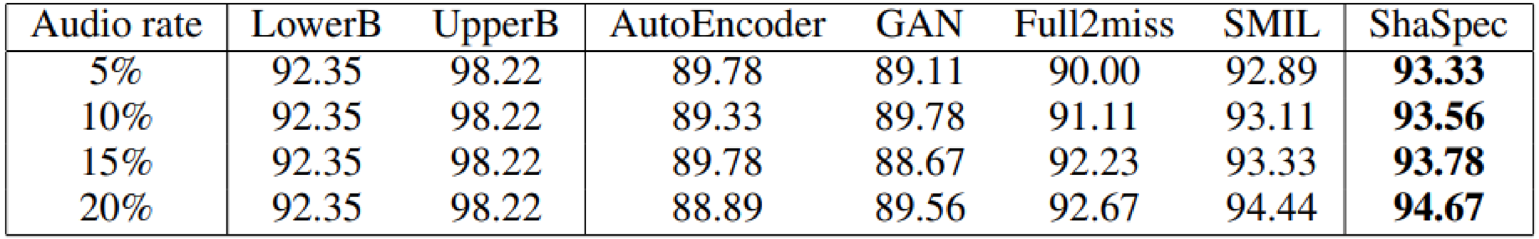
Table 1. Model performance comparison of **segmentation** Dice score (normalised to 100%) on BraTS2018 of **non-dedicated training**. ShaSpec and ShaSpec* are the proposed models, with ShaSpec* being the model with prediction smoothness enhancement. The best and second best results for each column within a certain type of tumour are in **red** and **blue**, respectively.



Modalities				Enhancing tumour				tumour Core				Whole tumour			
Fl	T1	T1c	T2	KD-Net	ACN	ShaSpec	ShaSpec*	KD-Net	ACN	ShaSpec	ShaSpec*	KD-Net	ACN	ShaSpec	ShaSpec*
●	○	○	○	40.99	42.77	**43.94**	**43.97**	65.97	67.72	**70.97**	**70.99**	85.14	87.30	**89.28**	**89.38**
○	●	○	○	39.87	41.52	**45.24**	**46.76**	70.02	**71.18**	70.28	**70.64**	77.28	79.34	**79.40**	**79.50**
○	○	●	○	75.32	**78.07**	75.91	**78.40**	81.89	84.18	**84.19**	**85.47**	76.79	**80.52**	80.43	**80.55**
○	○	○	●	39.04	42.98	**44.54**	**46.07**	66.01	67.94	**70.30**	**70.11**	82.32	85.55	**85.58**	**85.62**
Average				48.81	51.34	**52.41**	**53.80**	70.97	72.76	**73.92**	**74.30**	80.38	83.18	**83.67**	**83.76**

Table 2. Model performance comparison of **segmentation** Dice score (normalised to 100%) on BraTS2018 of **dedicated training**.

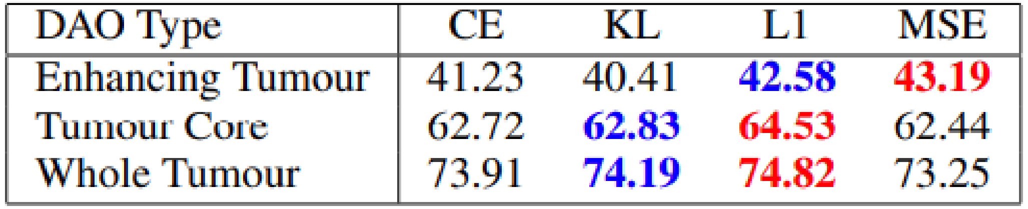
Experimental Results on Audio-Visual Classification Datasets



Audio rate	LowerB	UpperB	AutoEncoder	GAN	Full2miss	SMIL	ShaSpec
5%	92.35	98.22	89.78	89.11	90.00	92.89	**93.33**
10%	92.35	98.22	89.33	89.78	91.11	93.11	**93.56**
15%	92.35	98.22	89.78	88.67	92.23	93.33	**93.78**
20%	92.35	98.22	88.89	89.56	92.67	94.44	**94.67**

Table 3. Model performance comparison of classification accuracy of missing modality (by setting different available audio rates) on Audiovision-MNIST dataset. The lower bound (LowerB) is a LeNet [16] network trained with single modality (images only). The upper bound (UpperB) is a model trained with all data modalities (all images and audios). The best results for each row are bolded.

Analysis and Visualizations



DAO Type	CE	KL	L1	MSE
Enhancing Tumour	41.23	40.41	**42.58**	**43.19**
Tumour Core	62.72	**62.83**	**64.53**	62.44
Whole Tumour	73.91	**74.19**	**74.82**	73.25

Table 4. Model ablation of different distribution alignment objectives for non-dedicated training, where only T1 is available for testing on BraTS2018.

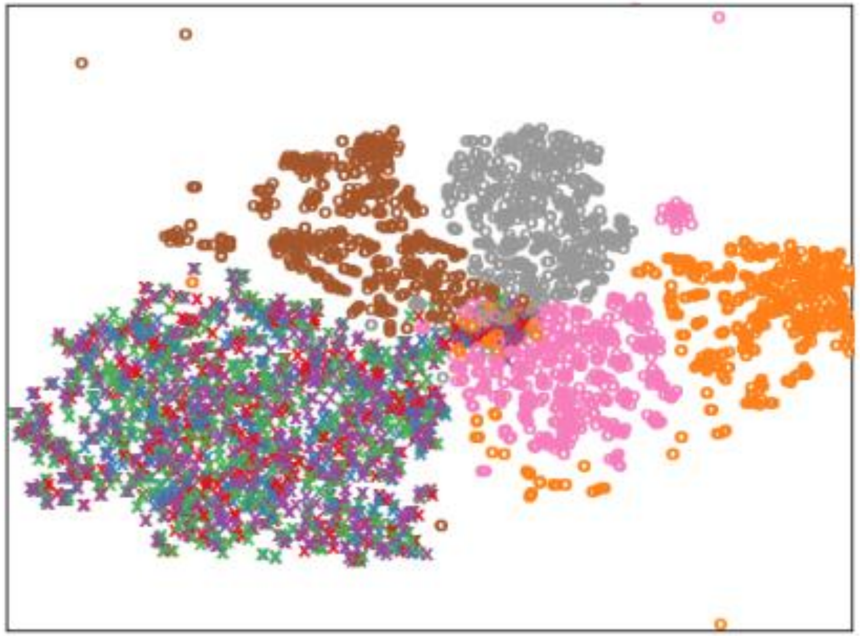


Figure 5. t-SNE visualisation of shared and specific features of four modalities from all training data on BraTS2018. The shared features of four modalities are presented by '**x**' in different colours, while the specific features of four modalities are presented by '**o**' in different colours.

Thank you!