Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Neuroimaging Cranial Pathology Delineation through Contrast-Specialized Specialist Routing

Grace Katuri

grace.katuri@accuityhealthcare.com

Accuity Healthcare USA https://orcid.org/0009-0002-3178-5560

Research Article

Keywords: Magnetic resonance imaging, Brain, Deep learning, Image segmentation, Glioma, Stroke, Multiple sclerosis

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11761457/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Purpose

Cranial MRI reaches the reading room in whatever contrast combination the scanner, site and protocol produced, and lesion appearance changes with contrast. A department that wants segmentation support must maintain a separate model for each contrast combination, a library that is hard to sustain in routine practice. We evaluated a single deep learning framework that replaces that library with contrast-specialised routing.

Materials and methods

The framework (OMNIA) keeps one specialist subnetwork for each MR contrast and fuses their predictions at every resolution level through a hierarchical gating network that assigns voxelwise weights. Progressive training protects each specialist's contrast-specific behaviour, and a soft allocation module routes incomplete contrast sets to the specialist channels under an availability prior. The study was retrospective, across nine collections totalling 6,336 annotated volumes, five contrasts and six pathology groups, with a healthy reference collection; six public collections were used for training and internal testing, and three institutional collections were held back for external testing.

Results

On the seen collections the framework reached a mean Dice of 0.8204, above every universal baseline (0.8007–0.8084) and close to dedicated per-combination pipelines (0.8250), as a single deployable model. On the unseen collections it reached 0.7951, the strongest result among single models. On multi-contrast BraTS 2021 input it scored 0.8334 against 0.8366 for fifteen dedicated models, at about one third of their storage (1.65 GB versus about 4.95 GB).

Conclusion

Contrast-specialised routing allows one deployable model to approach the accuracy of dedicated per-combination pipelines, on data and contrast combinations beyond those it was trained on.

1. Introduction

Volumetric delineation of cranial pathology on magnetic resonance imaging (MRI) underpins treatment planning, follow-up and response assessment across neuro-oncology and stroke care. Modern segmentation networks reach high accuracy when the input matches the contrast they were trained on, and self-configuring pipelines such as nnU-Net have made that accuracy routine to obtain [1]. The difficulty is that contrast is not fixed in clinical practice. A lesion's brightness, boundary definition and

internal structure differ between T1-weighted, contrast-enhanced T1-weighted, T2-weighted, FLAIR and diffusion-weighted images, so a network that performs well on one contrast may fail on another acquired from the same patient.

The conventional answer is to train and maintain a separate model for every contrast, and for every combination of contrasts a department might receive. That answer scales badly. Each model needs its own validation, version control, storage and monitoring, and the count grows combinatorially as protocols diversify across scanners and sites. A single universal model avoids the library but pays for it in accuracy: one encoder must absorb appearance differences that are large relative to the subtle cues separating lesion from healthy tissue, and performance settles below what dedicated models achieve.

Several directions have been explored. Promptable segmentation adapts at inference time [2], but a prompt describes where a lesion is rather than what it looks like, and small, diffuse or multifocal lesions are poorly served by point or box prompts. Missing-modality methods impute or share features across contrasts [3, 4], and contrast-agnostic synthesis can segment anatomy under any contrast [5], yet pathology delineation depends on exactly the contrast-specific signal these approaches smooth away. Transformer architectures [6, 7] and domain adaptation [8] improve robustness without changing the underlying premise that one representation should serve all contrasts. The BraTS benchmark illustrates the setting: tumours are presented in several contrasts precisely because no single contrast describes them fully [9].

This work takes a different position: treat contrast as an explicit axis of specialisation inside one model. We present OMNIA, a framework in which each contrast is handled by its own specialist subnetwork, and a hierarchical gating network combines the specialists' predictions at every resolution level with weights that vary voxel by voxel. Progressive training keeps specialists faithful to their contrast while the fused output learns to arbitrate between them, and a soft allocation module routes incomplete contrast sets onto the specialist channels, so the same model serves a lone FLAIR study and a full four-contrast protocol. The pipeline builds on established open infrastructure for registration, anatomical processing and training [10–13]. We evaluate the framework retrospectively on nine collections spanning five contrasts and six pathology groups, with a healthy reference collection, with three institutional collections reserved for external testing, and compare it with universal models and with dedicated per-combination pipelines.

2. Materials and methods

2.1. Study design and data collections

The study is a retrospective analysis of existing, previously annotated MRI collections; no new imaging was acquired. Nine collections were assembled, totalling 6,336 volumes across five contrasts (T1w, T1ce, T2w, FLAIR, DWI) and six pathology groups, with a healthy reference collection (Table 1). Six are public benchmarks: ATLAS 2.0 (chronic stroke), ISLES 2022 (acute ischaemic stroke), BraTS 2021

(glioma), WMH 2017 (white-matter hyperintensities), MSSEG (multiple sclerosis) and OASIS (healthy reference scans). The remaining three (Tumour1, Tumour2, WMHmix) are institutional collections covering glioma and mixed white-matter disease. Every volume was rigidly registered to a common template, skull-stripped, and cropped or resampled to a shared volume size and isotropic voxel spacing. The public collections formed the training pool and the internal test sets; the institutional collections were never used in training and serve as external test data.

Table 1
Collections analysed. Six public benchmarks formed the training pool and internal test sets; the three institutional collections were reserved for external testing.

Collection	Task	Pathology focus	Train	Test	Contrasts
ATLAS 2.0	Stroke lesion	Chronic stroke	545	110	T1w
ISLES 2022	Stroke lesion	Ischaemic stroke	150	100	T1w, FLAIR, DWI
BraTS 2021	Tumour	Glioma	4168	836	T1w, T1ce, T2w, FLAIR
WMH 2017	WM lesion	White-matter hyperintensity	60	36	T1w, FLAIR
MSSEG	MS lesion	Multiple sclerosis	20	24	T1w, T2w, FLAIR
OASIS	Healthy ref.	None (reference)	27	20	T1w
Tumour1	Tumour	Glioma	40	40	T1w, T1ce, T2w, FLAIR
Tumour2	Tumour	Glioma	40	40	T1w, T1ce, T2w, FLAIR
WMHmix	WM lesion	Mixed WM disease	40	40	T1w, FLAIR

2.2. Formulation

Consider N contrasts in total. Each specialist learns a contrast-specific delineation mapping from its own contrast's imaging domain to the shared label space of lesion masks; the framework as a whole realises a mapping that accepts any contrast, or any combination of contrasts, and returns one mask:

$f : X \rightarrow Y$ (1)

where X is an input volume and Y its delineation. For multi-contrast input, the volume set is written

$\tilde{X} = (X_1, \ldots, X_N)^T$ (2)

and availability is encoded by a binary vector c, whose entry for a contrast is 1 when that contrast is present and 0 otherwise. The task is to delineate pathology under whatever availability pattern a given examination presents.

2.3. Contrast-specialised routing

The framework holds one specialist subnetwork per contrast, each a three-dimensional encoder–decoder that outputs delineation features at several resolutions. A hierarchical gating network receives the specialists' feature maps at every stage and produces, for each stage, one weighting map per specialist; the stage output is the weighted sum (Fig. 1):

$o^{l} = \Sigma_{i=1}^{N} e_i^{l} \odot g_i^{l}$ (3)

where ⊙ is elementwise multiplication. The weights are normalised across specialists at each voxel by a softmax, so the gating decides locally which contrast's evidence to trust. Every stage output is supervised with a deep-supervision loss, and the stage outputs are combined for the final delineation.

Jointly trained specialists tend to degenerate: the gating learns to lean on one or two dominant contrasts, the neglected specialists stop improving, and the bank quietly collapses into a universal model with extra parameters. The failure is most likely exactly where specialisation matters most, on contrasts whose lesion appearance is subtle.

Training therefore follows a progressive objective that changes the balance of losses over time:

$L_{total} = f_{ep} \cdot L_{specialists} + (1 - f_{ep}) \cdot L_{fused}, f_{ep} = (1 - ep/EP)^2$ (4)

where the first term sums the losses of the individual specialists on their matched contrasts, the second term is the loss of the combined output, ep is the current epoch and EP the total number of epochs. Early in training the weighting factor is close to one and specialists are forced to become competent on their own contrast; as it decays (Fig. 2), the fused objective takes over and the gating learns to combine already-competent specialists. A reduced learning rate is used while this objective is active.

Each specialist is penalised on its matched contrast and on the multi-contrast input:

$L_{specialists} = \Sigma_i [\, L(e_i(X_i), Y) + \lambda_i L(e_i(\tilde{X}), Y) \,]$ (5)

$L_{fused} = \Sigma_l w_l L(o^{l}, Y)$ (6)

with L a composite segmentation loss, the λ terms weighting each specialist's multi-contrast penalty and the w terms weighting the stage outputs of deep supervision. Penalising specialists on multi-contrast input exposes every specialist to combined evidence, which the gating later exploits.

2.4. Soft allocation for incomplete contrast sets

A department cannot guarantee that every examination arrives with the contrasts the specialists expect. The soft allocation module (Fig. 3) closes that gap. An allocation subsystem maps the availability vector c to a matrix D of routing weights; a constrained version D* is then formed by zeroing every entry that draws on an absent contrast and renormalising each row:

$D = softmax_{rows}(f_D(c)),\ D^*_{ij} = D_{ij}\ if\ c_j = 1,\ else\ 0$ (7)

so no weight is ever assigned to a contrast that is absent. The routed channel set is

$S = D^* \tilde{X}$ (8)

and each specialist receives the routed mixture in place of a raw contrast. When the input already matches a specialist's contrast, the routing reduces to the identity and behaviour is unchanged. Two regimes arise. If at least one expected contrast is present, the available contrasts are routed and the corresponding specialists activate. If none of the expected contrasts is present, every expert output is replaced by a dataset-level prior estimate, so the framework degrades gracefully instead of failing.

Allocation is trained with contrast dropout: during multi-contrast training, contrasts are randomly removed with probability 0.5 under the objective

$L_{multi} = \Sigma_{c \in C} L(f(\tilde{X}, c), Y)$ (9)

where C enumerates availability patterns. The availability vector is also concatenated to the bottleneck of the gating pathway, giving the fusion an explicit prior over which specialists can be trusted for the examination at hand.

2.5. Training and implementation

Training ran in two stages on a single NVIDIA A100: specialists were first trained independently on their own contrasts, then the gating network was trained with the specialists fine-tuned under the progressive objective, for 800 epochs in total. Specialists and gating share the same volumetric backbone design; deep supervision used its default stage weights throughout.

2.6. Comparators and evaluation

Six comparators were evaluated under identical data conditions: (i) a single universal model trained on all contrasts pooled; (ii) a multi-dataset model trained across collections; (iii) a context-prior universal model using dataset and contrast conditioning; (iv) a residual-encoder variant of the universal model; (v) a promptable three-dimensional medical segmenter [2]; and (vi) dedicated pipelines trained separately for each dataset–contrast combination, included as the accuracy reference rather than as a deployable option. Accuracy is reported as Dice similarity coefficient at image, task and dataset level, with Hausdorff distance for boundary agreement. Deployment cost is summarised by the number of models a service must maintain, training memory, disk storage and training time.

3. Results

3.1. Seen collections

On the collections used in training (Table 2), the proposed framework achieved the highest image-level Dice of any single model, 0.8204, with task-level Dice 0.7026 and dataset-level Dice 0.6938; the universal baselines ranged from 0.8007 to 0.8084 at image level. The promptable segmenter was weakest

(0.7322), consistent with the difficulty of prompting small or diffuse lesions. The dedicated pipelines remained the accuracy reference at 0.8250, but they require fourteen separately trained models; the framework is one. Its training memory (38.2 GB) exceeds the 7.5 GB of a single universal model, as expected for a specialist bank, while staying far below the combined footprint of the dedicated set. Hausdorff distance was competitive at 24.51, against 23.80 for the dedicated pipelines and 25.99–31.03 for the universal baselines. Gains over the universal models were largest on T1w input, and performance was consistent across glioma, stroke and white-matter pathology. In the learned feature space, the routed multi-contrast representation separated lesion classes more cleanly than the representation of any universal baseline.

Table 2
Performance on the seen collections. Dice is reported at image, task and dataset level; HD is Hausdorff distance. Memory and hours refer to training.

Method	Dice_img	Dice_task	Dice_data	HD	Mem (GB)	Hours
Universal (single model)	0.8007	0.6749	0.6646	31.03	7.5	9.3
Universal + context prior	0.8063	0.6899	0.6729	28.09	7.5	9.1
Multi-dataset model	0.8048	0.6916	0.6831	27.58	7.5	9.4
Dedicated (per combination)	0.8250	0.7100	0.7000	23.80	7.5×14	9.1×14
Residual-encoder universal	0.8084	0.6911	0.6815	25.99	7.5	10.2
Promptable 3D segmenter	0.7322	0.6412	0.6231	30.45	7.5	9.8
Proposed (OMNIA)	0.8204	0.7026	0.6938	24.51	38.2	46.6

3.2. External testing on unseen collections

On the three institutional collections, none of which contributed to training, the framework reached a mean Dice of 0.7951 (Tumour1 0.8545, Tumour2 0.8293, WMHmix 0.7015), ahead of every universal baseline on every collection (Table 3). The strongest universal model reached 0.7803. Notably, the framework also exceeded the dedicated pipelines on both tumour collections, even though those pipelines had been trained with access to them; the dedicated pipelines kept their advantage only on WMHmix (0.7508), whose mixed white-matter subtypes vary widely in appearance.

Table 3
External testing on the three unseen institutional collections (Dice). Dedicated pipelines were trained with access to these collections; the proposed framework was not.

Method	Tumour1	Tumour2	WMHmix	Mean
Universal (single model)	0.8012	0.7885	0.6740	0.7546
Universal + context prior	0.8298	0.7921	0.6815	0.7678
Multi-dataset model	0.8310	0.8022	0.6902	0.7745
Dedicated (per combination)	0.8450	0.8228	0.7508	0.8061
Residual-encoder universal	0.8355	0.8150	0.6905	0.7803
Promptable 3D segmenter	0.7845	0.7602	0.6511	0.7319
Proposed (OMNIA)	0.8545	0.8293	0.7015	0.7951

3.3. Ablation

Removing components degraded performance in a consistent order (Table 4). The full framework scored 0.8204 on seen and 0.7951 on unseen collections. Dropping the progressive objective cost accuracy on both (0.8142 and 0.7889), and removing soft allocation as well left 0.8117 and 0.7912. With neither component, performance fell to 0.8027 and 0.7814, close to the universal baselines. Inspection of the gating weights confirmed the mechanism: without the progressive schedule, some specialists fall silent and whole contrasts are carried by a single expert, whereas the scheduled model keeps all specialists active, with T1w and T1ce frequently co-activated, as expected from their shared acquisition physics.

Table 4
Ablation of the two routing components, on seen and unseen collections (Dice).

Variant	Seen	Unseen
Full framework	0.8204	0.7951
Without progressive training	0.8142	0.7889
Without progressive training or soft allocation	0.8117	0.7912
Neither component	0.8027	0.7814

3.4. Multi-contrast examinations

For examinations that do arrive with several contrasts, the comparison is against a library: the four BraTS contrasts admit 15 non-empty combinations, and the dedicated approach trains a model for each (Table 5). The unified framework matched that library closely, with mean Dice 0.8334 against 0.8366, a gap of about a third of a percentage point, while reducing training memory from over 112.5 GB to 30.56

GB and storage from approximately 4.95 GB to 1.65 GB. Variants that shared a decoder across specialists were cheaper still but slightly less accurate; the full specialist bank gave the best delineation.

Table 5
Multi-contrast BraTS 2021 evaluation. The dedicated approach trains one model per non-empty contrast combination (15); the proposed framework is a single model.

Approach	Models	Mean Dice	Train mem. (GB)	Storage (GB)
Dedicated per combination	15	0.8366	112.5	≈ 4.95
Shared decoder across specialists	1	0.8312	22.41	1.12
Proposed (full specialist bank)	1	0.8334	30.56	1.65

4. Discussion

The results support a simple reading: for cranial pathology, contrast is not noise to be normalised away but structure to be modelled. A universal encoder must spend capacity reconciling appearance differences between contrasts before it can attend to the lesion; specialists start from the appearance of one contrast and spend their capacity on the lesion itself. The gating then recovers, voxel by voxel, the benefit of having several contrasts available, without surrendering the benefit of specialisation.

For a radiology service, the practical consequence is operational. One model is deployed, validated and monitored; no staff member selects a model at the console, and no examination fails because its protocol was unusual. External testing matters here: the collections the framework had never seen are the closest analogue in this study to a new site's data, and they are where the single-model advantage over universal baselines was clearest. The multi-contrast analysis makes the same point from the other direction, since the unified model held its accuracy as the combination count grew, where the dedicated library grows combinatorially.

The margins were not uniform. They were largest where lesion appearance is most contrast-dependent and smallest on the mixed white-matter collection, whose subtypes differ from one another almost as much as contrasts do. That pattern is informative rather than discouraging: it identifies heterogeneity within a pathology label, not contrast handling, as the next bottleneck.

Several limitations bound these findings. Each volume carries a single pathology annotation, so lesions that co-occur — a tumour within chronic white-matter disease, for instance — are outside what the framework was trained or tested to separate; multi-pathology annotation and lesion-aware augmentation are the natural next steps. Evaluation is confined to the brain; whether routing transfers to other organs and modalities is untested. The specialist bank uses more memory than a single universal model, though far less than the dedicated libraries it replaces. Finally, the allocation assumes the availability pattern of an examination is known, which holds for DICOM-routed clinical data but should be verified when the framework is embedded in a new pipeline.

Future work will extend the routing idea in three directions: training with multi-pathology annotations so co-occurring disease can be delineated jointly; carrying the same specialist-and-gating design to other modality families; and calibrating the gating weights so that their voxelwise pattern can be shown to the reporting radiologist as an interpretable summary of which contrast drove each part of the delineation.

5. Conclusion

Contrast-specialised routing replaces a library of per-contrast and per-combination models with one deployable framework. Across nine collections it exceeded universal baselines on seen data, generalised best among single models to unseen collections, and matched a fifteen-model dedicated library on multi-contrast input at a third of the storage. For departments whose MRI arrives in whatever contrast combination the day produced, that combination of accuracy and simplicity is the point.

Declarations

Competing interests.

The author declares no competing interests.

Ethics approval.

This study is a retrospective analysis of existing medical imaging collections. No participants were recruited for this study, no new data were collected, and no identifiable patient information was available to the author at any stage. Six of the nine collections are public benchmark datasets. Ethical approval was not required for this retrospective secondary analysis, and no consent was required, as the study involved no direct contact with participants and no identifiable information.

Disclosure.

This manuscript was prepared under the direct direction and supervision of the author, with drafting and editorial assistance used during its preparation. The author reviewed and edited the material and takes full responsibility for the content of this publication.

Funding.

This research received no external funding.

Author contributions.

The sole author carried out all aspects of the study and the preparation of this manuscript.

Data availability.

Six of the nine collections analysed in this study are public datasets, available from their respective repositories under their published data use terms: ATLAS 2.0, ISLES 2022, BraTS 2021, WMH 2017, MSSEG and OASIS. The remaining three collections (Tumour1, Tumour2 and WMHmix) are institutional research collections and are not publicly available. No new datasets were generated by this study.

References

1. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH (2021) nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18:203–211. https://doi.org/10.1038/s41592-020-01008-z
2. Ma J, He Y, Li F, Han L, You C, Wang B (2024) Segment anything in medical images. Nat Commun 15:654. https://doi.org/10.1038/s41467-024-44824-z
3. Konwer A, Chen C, Prasanna P (2023) MagNET: modality-agnostic network for brain tumor segmentation and characterization with missing modalities. Machine Learning in Medical Imaging (MLMI 2023), Lecture Notes in Computer Science, vol 14349. Springer, Cham, pp 361–371. https://doi.org/10.1007/978-3-031-45673-2_36
4. Havaei M, Guizard N, Chapados N, Bengio Y (2016) HeMIS: hetero-modal image segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer, Cham, pp 469–477. https://doi.org/10.1007/978-3-319-46723-8_54
5. Billot B, Greve DN, Puonti O, Thielscher A, Van Leemput K, Fischl B, Dalca AV, Iglesias JE (2023) SynthSeg: segmentation of brain MRI scans of any contrast and resolution without retraining, Med. Image Anal 86:102789. https://doi.org/10.1016/j.media.2023.102789
6. Hatamizadeh A, Tang Y, Nath V, Yang D, Myronenko A, Landman B, Roth HR, Xu D (2022) UNETR: transformers for 3D medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), IEEE, pp. 1748–1758. https://doi.org/10.1109/WACV51458.2022.00181
7. Hatamizadeh A, Nath V, Tang Y, Yang D, Roth HR, Xu D, Swin UNETR (2022) Swin transformers for semantic segmentation of brain tumors in MRI images, in: International Workshop on Brainlesion (BrainLes – MICCAI), Lecture Notes in Computer Science, vol. 12962, Springer, Cham, pp. 272–284. https://doi.org/10.1007/978-3-031-08999-2_22
8. Kamnitsas K, Baumgartner C, Ledig C, Newcombe V, Simpson J, Kane A, Menon D, Nori A, Criminisi A, Rueckert D, Glocker B (2017) Unsupervised domain adaptation in brain lesion segmentation with adversarial networks. Information Processing in Medical Imaging (IPMI), Lecture Notes in Computer Science, vol 10265. Springer, Cham, pp 597–609. https://doi.org/10.1007/978-3-319-59050-9_47

9. Baid U, Ghodasara S, Mohan S, Bilello M, Calabrese E, Colak E, Farahani K, Kalpathy-Cramer J, Kitamura FC, Pati S et al (2021) The RSNA-ASNR-MICCAI BraTS 2021 benchmark on brain tumor segmentation and radiogenomic classification. arXiv preprint. https://doi.org/10.48550/arXiv.2107.02314. arXiv:2107.02314
10. Tustison NJ, Cook PA, Holbrook AJ, Johnson HJ, Muschelli J, Devenyi GA, Duda JT, Das SR, Cullen NC, Gillen DL, Yassa MA, Stone JR, Gee JC, Avants BB (2021) The ANTsX ecosystem for quantitative biological and medical imaging. Sci Rep 11:9068. https://doi.org/10.1038/s41598-021-87564-6
11. Wasserthal J, Breit H-C, Meyer MT, Pradella M, Hinck D, Sauter AW, Heye T, Boll DT, Cyriac J, Yang S, Bach M, Segeroth M (2023) TotalSegmentator: robust segmentation of 104 anatomic structures in CT images. Radiol Artif Intell 5:e230024. https://doi.org/10.1148/ryai.230024
12. Henschel L, Conjeti S, Estrada S, Diers K, Fischl B, Reuter M (2020) FastSurfer – a fast and accurate deep learning based neuroimaging pipeline. NeuroImage 219:117012. https://doi.org/10.1016/j.neuroimage.2020.117012
13. Cardoso MJ, Li W, Brown R, Ma N, Kerfoot E, Wang Y, Murrey B et al (2022) MONAI: an open-source framework for deep learning in healthcare, arXiv preprint arXiv:2211.02701. https://doi.org/10.48550/arXiv.2211.02701

Figures

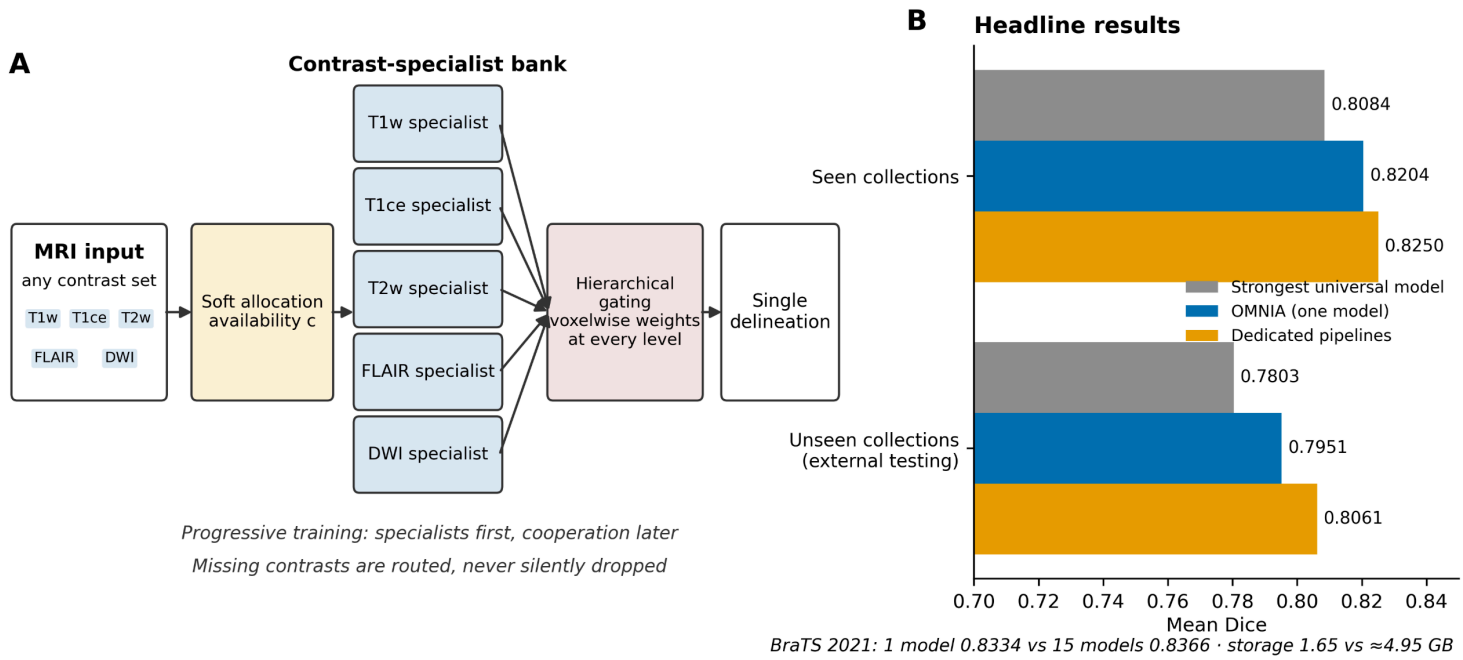


Figure 1

Overview of the framework and headline results. (A) An examination in any available contrast set is routed by soft allocation to a bank of contrast specialists; hierarchical gating merges their multilevel predictions with voxelwise weights into a single delineation. (B) Mean Dice on seen collections and on

external testing in three unseen collections, for the strongest universal baseline, the proposed framework and dedicated per-combination pipelines.

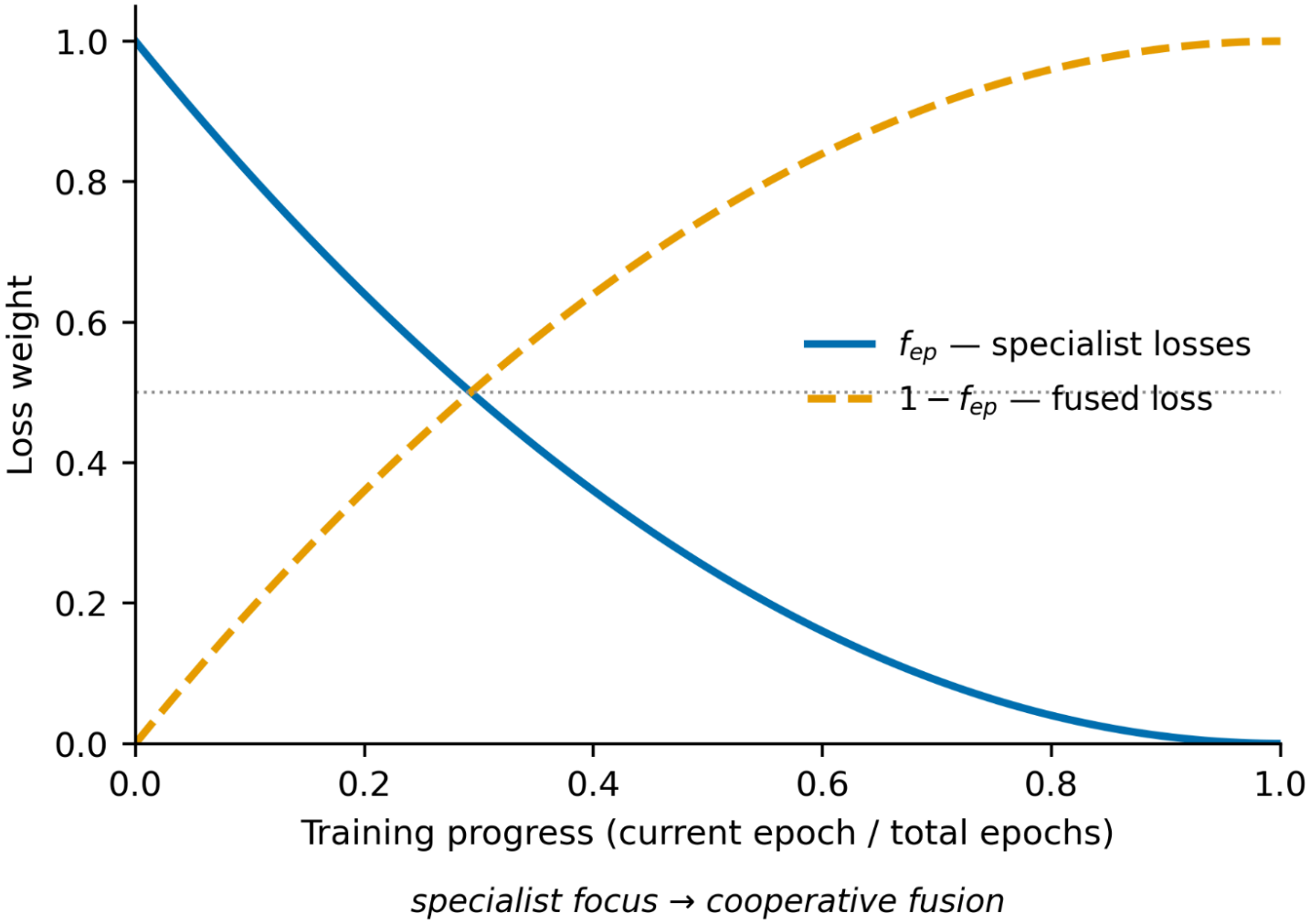


Figure 2

The progressive training schedule. The specialist weight f_ep decays across training while the fused objective takes over, protecting specialisation early and encouraging cooperation late.

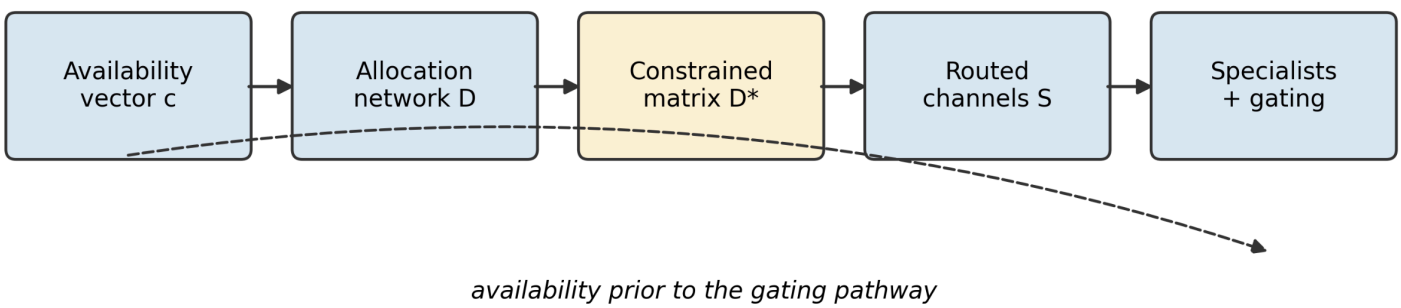


Figure 3

Soft allocation for incomplete contrast sets. The availability vector c drives an allocation network whose constrained matrix D* routes the available contrasts onto the specialist channels; c is also passed to the gating pathway as a prior.