RBT-Faz: Register-Bottleneck 3D Swin-UNETR with Clinical Text Alignment for Ordinal Fazekas Grading from FLAIR MRI

Farzan Niknejad Mazandarani[1][0000−0001−8599−729X], Kai La Jennifer Tsang[1][0009−0007−5243−6251], Nicholas Simard[1][0000−0003−2624−8728], Mumu Aktar[1][0000−0002−5718−7238], and April Khademi[1,2,3,4,5][0000−0002−4932−0385]

[1] Department of Electrical, Computer, and Biomedical Engineering, Toronto Metropolitan University, Toronto, ON, Canada
{fniknejad,tsangk9,nicholas.simard,mumu.aktar,akhademi}@torontomu.ca
[2] Institute for Biomedical Engineering, Science Tech (iBEST), Toronto, ON, Canada
[3] Vector Institute for Artificial Intelligence, Toronto, ON, Canada
[4] Institute of Medical Science, University of Toronto, Toronto, ON, Canada
[5] Department of Medical Imaging, University of Toronto, Toronto, ON, Canada

Abstract. White matter hyperintensities (WMHs) are commonly scored using the ordinal Fazekas scale, but manual assessment is subjective and laborious to obtain. We propose RBT-Faz, a direct 3D fluid-attenuated inversion recovery (FLAIR) MRI framework for subject-level periventricular hyperintensity (PVH) and deep white matter hyperintensity (DWMH) Fazekas scoring. RBT-Faz uses a 3D Swin-UNETR encoder and introduces a final-layer Firewall Attention Block, where learnable register tokens aggregate global WMH severity context from localized patch tokens through a register bottleneck. This design preserves lesion-sensitive spatial features while forming compact subject-level severity representations. To improve ordinal consistency and clinical grounding, the visual embedding is aligned with ClinicalBERT-encoded ground-truth Fazekas grade descriptors, and a lesion-aware auxiliary head predicts WMH burden, periventricular/deep spatial distribution, and lesion count during training only. Experiments on Canadian Consortium on Neurodegeneration in Aging (CCNA) subjects showed that RBT-Faz outperformed volumetric classifiers and segmentation-based rule grading, achieving Macro-F1/quadratic weighted kappa (QWK) of 0.786/0.886 for PVH and 0.825/0.914 for DWMH. Grad-CAM and descriptor analyses further supported clinically meaningful model behavior. The anonymized code repository is available at: `https://github.com/IAMLAB-Ryerson/RBT-Faz`.

Keywords: Fazekas grading · White matter hyperintensities · Ordinal representation learning.

1 Introduction

White matter hyperintensities (WMHs) are common radiological markers of cerebral small vessel disease and are routinely evaluated on fluid-attenuated inversion

recovery (FLAIR) magnetic resonance imaging (MRI) [16, 13]. Their severity is clinically associated with aging, cognitive decline, stroke risk, and neurodegenerative progression [1, 3]. In routine neuroradiology workflows, WMH burden is commonly summarized using the Fazekas scale, an ordinal score from 0 to 3 that separately describes periventricular hyperintensity (PVH) and deep white matter hyperintensity (DWMH) involvement. Although the Fazekas scale is easy to perform and widely used in clinical practice, manual grading remains subjective, time-consuming, and affected by inter-rater variability [19]. In addition, its coarse ordinal categories may limit sensitivity to subtle longitudinal changes, which can make tracking disease progression challenging in clinical follow-up. Automated Fazekas classification is therefore important for improving reproducibility and scalability in WMH severity assessment.

Existing automated Fazekas grading methods can be broadly divided into segmentation-based and direct prediction approaches. The most common strategy first segments WMH lesions from FLAIR MRI and then assigns Fazekas grades using either predefined clinical rules, such as lesion volume, diameter, or anatomical location, or a subsequent learned classifier trained on lesion masks or derived features [10, 12, 19]. These methods are interpretable and can match radiologist agreement, but depend heavily on segmentation quality and may suffer error propagation or limited threshold generalization. More recent direct prediction methods estimate Fazekas scores from MRI without explicit lesion segmentation, reducing annotation needs and simplifying inference [20, 15]. However, direct Fazekas grading is challenging because it requires models to infer global disease severity from subject-level labels while also capturing subtle local WMH patterns, such as periventricular caps and pencil-thin linings.

To address this challenge, we propose *RBT-Faz*, a register-bottleneck 3D Swin-UNETR framework for direct ordinal Fazekas scoring from FLAIR MRI volumes. The input is only the MRI volume, which is partitioned into non-overlapping 3D patches and processed by a hierarchical 3D Swin-UNETR encoder [7]. At the final encoder layer, we introduce a *Firewall Attention Block (FAB)*, where learnable register tokens serve as semantic query tokens. FAB aggregates global WMH severity context from localized visual patch tokens through unidirectional patch-to-register attention, followed by register-to-register refinement. This design mediates long-range information flow through compact register tokens rather than unrestricted global patch-to-patch mixing, helping preserve local WMH-sensitive evidence while forming subject-level disease prototypes.

In addition, RBT-Faz formulates Fazekas grading as an ordinal representation learning problem. A text-manifold alignment module aligns the visual embedding with ClinicalBERT-encoded ground-truth Fazekas grade descriptors for the target subscale [8], including PVH patterns such as caps, pencil-thin lining, smooth halo, and irregular periventricular extension, as well as DWMH patterns such as punctate foci, beginning confluence, and large confluent hyperintensities. We further introduce a lesion-aware auxiliary head during training to encourage the shared representation to encode WMH burden, periventricular/deep spatial

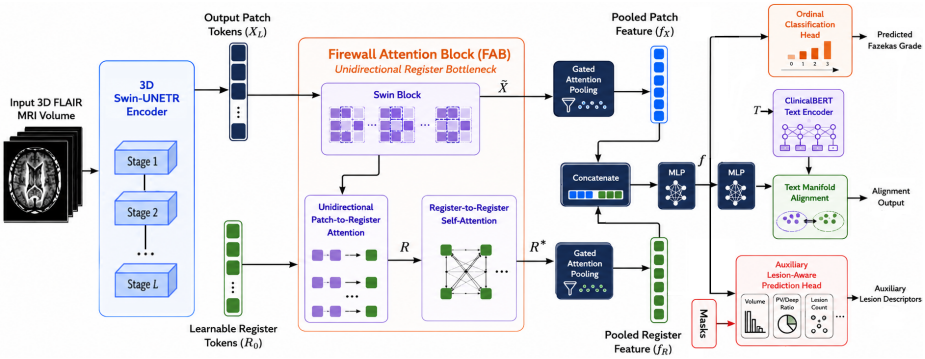


Fig. 1: Schematic overview of RBT-Faz for direct ordinal Fazekas grading from 3D FLAIR MRI. During inference, only the ordinal classification head is used.

distribution, and lesion count. Importantly, this auxiliary branch is used only for supervision and does not require lesion masks or anatomical priors as model inputs during inference.

The remainder of this paper is organized as follows: Sect. 2 details the proposed RBT-Faz methodology; Sect. 3 presents the experimental validation, comparative results, and clinical interpretability analyses; and Sect. 4 concludes the paper.

2 Methodology

Problem setup. Given a 3D FLAIR MRI volume $\mathbf{V} \in \mathbb{R}^{H\times W\times D\times C}$ (height H, width W, depth D, and channel dimension C), the goal is to predict an ordinal Fazekas subscore $y^s \in \{0,1,2,3\}$ for target subscale $s \in \{\mathrm{pv},\mathrm{dw}\}$, corresponding to PVH or DWMH. We train a separate model for each subscale; when needed, a total Fazekas summary can be obtained post hoc by combining the predicted PVH and DWMH subscores. The input is restricted to the MRI volume only, without hyperintensity masks, anatomical priors, or predefined clinical features during inference. The volume is partitioned into non-overlapping $2 \times 2 \times 2$ 3D patches, flattened, and linearly embedded into visual patch tokens $\mathbf{X}_0 \in \mathbb{R}^{N\times d}$, where N is the number of patches and d is the token dimension.

Hierarchical 3D Swin-UNETR encoder. The patch tokens are processed by a hierarchical 3D Swin-UNETR encoder [7]. The encoder uses window-based and shifted-window self-attention to extract multiscale volumetric features, while patch merging progressively reduces spatial resolution and increases feature abstraction. Since Fazekas grading is a subject-level ordinal classification task rather than dense segmentation, the decoder is removed and the final encoder representation is used for classification. This preserves the strength of Swin-UNETR in modeling local-to-global volumetric context while adapting the architecture to image-level WMH severity grading. Removing the decoder and dis-

pensing with lesion segmentation at inference also reduces inference-time computation and model footprint, which may benefit deployment in large-scale clinical or research workflows. The overall schematic of the proposed RBT-Faz framework is shown in Fig. 1.

Firewall Attention Block. Inspired by register-mediated attention mechanisms [22], we introduce the Firewall Attention Block (FAB) at the final encoder layer to aggregate global WMH severity context while restricting excessive global mixing among dense spatial tokens. Let $\mathbf{X}_L \in \mathbb{R}^{N_L \times d}$ denote the final-stage visual patch tokens. A set of N_r learnable register tokens $\mathbf{R}_0 \in \mathbb{R}^{N_r \times d}$ is introduced at this stage. The patch tokens are first locally refined by the final Swin block, producing $\tilde{\mathbf{X}}$. The register tokens then aggregate information from the refined patch tokens through unidirectional patch-to-register attention. Specifically, queries are generated from the register tokens, while keys and values are generated from the locally refined patch tokens:

$$\mathbf{Q}_R = \mathbf{R}_0 \mathbf{W}_Q^R, \qquad \mathbf{K}_X = \tilde{\mathbf{X}} \mathbf{W}_K^X, \qquad \mathbf{V}_X = \tilde{\mathbf{X}} \mathbf{W}_V^X. \tag{1}$$

The register update is then computed as

$$\tilde{\mathbf{R}} = \mathrm{softmax}\left(\frac{\mathbf{Q}_R \mathbf{K}_X^\top}{\sqrt{d}}\right) \mathbf{V}_X. \tag{2}$$

In this formulation, $\mathbf{Q}_R$ denotes register-derived queries, whereas $\mathbf{K}_X$ and $\mathbf{V}_X$ denote patch-derived keys and values. Therefore, the attention map has size $N_r \times N_L$ and updates only the register stream. FAB imposes an asymmetric information pathway in which localized patch evidence is summarized into a compact set of register tokens, while the dense patch tokens are not globally updated by the registers within the same block. This constraint prevents unrestricted global patch-to-patch interaction and helps preserve localized WMH-sensitive features. The updated register tokens are subsequently refined using register-to-register self-attention, producing $\mathbf{R}^*$ as compact global semantic disease prototypes. The output of FAB consists of two complementary representations, namely localized patch tokens $\tilde{\mathbf{X}}$ and global register tokens $\mathbf{R}^*$.

Gated attention pooling and fusion. The patch and register streams are pooled separately using gated attention pooling rather than average pooling. For a token set $\mathbf{U} = \{\mathbf{u}_i\}_{i=1}^M$, a normalized learned gate $\alpha_i = \mathrm{softmax}_i\big(\mathbf{w}^\top \tanh(\mathbf{W}\mathbf{u}_i)\big)$ controls the contribution of token $\mathbf{u}_i$ to the pooled representation $\mathbf{f}_U = \sum_{i=1}^M \alpha_i \mathbf{u}_i$. Applying this operation to $\mathbf{R}^*$ and $\tilde{\mathbf{X}}$ gives $\mathbf{f}_R$ and $\mathbf{f}_X$, which are concatenated and projected as $\mathbf{f} = \mathrm{MLP}_{fuse}([\mathbf{f}_R; \mathbf{f}_X])$.

Ordinal Fazekas prediction. The fused representation $\mathbf{f}$ is passed to a classification head to estimate the probability of each ordinal grade, $\mathbf{p}^s = \mathrm{softmax}(\mathbf{W}_{cls}^s \mathbf{f} + \mathbf{b}_{cls}^s)$, where $\mathbf{p}^s \in \mathbb{R}^4$. The predicted subscore is $\hat{y}^s = \arg\max_k p_k^s$. To handle class imbalance, the classification head is trained with focal loss, $\mathcal{L}_{cls}^s = -\alpha_y (1 - p_y^s)^\gamma \log(p_y^s)$. Since Fazekas grading is ordinal, we additionally use an ordinal consistency loss based on the discrepancy between cumulative predicted and ground-truth distributions. Let $\mathbf{e}_{y^s}$ denote the one-hot encoding of

the ground-truth grade y^s and let $K = 4$ be the number of grades. The ordinal loss is defined as

$$\mathcal{L}^s_{ord} = \frac{1}{K-1} \sum_{k=0}^{K-2} \left(\sum_{j=0}^{k} p^s_j - \sum_{j=0}^{k} e_{y^s,j} \right)^2 . \tag{3}$$

This penalizes larger grading errors more strongly than adjacent-grade errors.

Text-manifold alignment. To align the image representation with the clinical meaning of each ordinal grade, we construct a text manifold from the ground-truth Fazekas visual rating descriptors. These descriptors define grade-specific imaging patterns for PVH and DWMH and are used as clinical text anchors. For PVH, the anchors are: T^{pv}_0 = absence, T^{pv}_1 = caps or pencil-thin lining, T^{pv}_2 = smooth halo, and T^{pv}_3 = irregular PVH extending into the deep white matter. For DWMH, the anchors are: T^{dw}_0 = absence, T^{dw}_1 = punctate foci, T^{dw}_2 = beginning confluence of foci, and T^{dw}_3 = large confluent areas. A frozen ClinicalBERT encoder [8] maps each target-specific grade descriptor into a normalized text embedding, $\mathbf{t}^s_k = \text{norm}(\text{ClinicalBERT}(T^s_k))$, where $k \in \{0, 1, 2, 3\}$. In parallel, a visual projection head maps the fused visual feature into the same embedding space, $\mathbf{z}^s_v = \text{norm}(\text{MLP}^s_{proj}(\mathbf{f}))$. The alignment objective is:

$$\mathcal{L}^s_{align} = -\log \frac{\exp((\mathbf{z}^s_v)^\top \mathbf{t}^s_{y^s}/\tau)}{\sum_{k=0}^{3} \exp((\mathbf{z}^s_v)^\top \mathbf{t}^s_k/\tau)}, \tag{4}$$

where τ is the temperature parameter, set to 0.07 in all experiments. This loss encourages the visual representation to align with the correct Fazekas grade descriptor for the selected subscale.

Auxiliary lesion-aware head and training objective. To improve anatomical grounding, we introduce an auxiliary lesion-aware head during training. The auxiliary targets are not manually provided as additional model inputs; instead, they are computed offline from pseudo-WMH and Ventricle masks generated by a pretrained nnU-Net model [9]. Given the predicted WMH mask $\hat{\mathbf{M}}$ and brain mask $\mathbf{B}$, we quantify three lesion-level descriptors: WMH volume ratio, periventricular-versus-deep WMH involvement, and lesion count. The WMH volume ratio is computed as $a_{vol} = |\hat{\mathbf{M}}|/|\mathbf{B}|$. To estimate regional involvement, similar to [4], we define a periventricular region using a distance map from the lateral ventricles and assign WMH voxels within radius $r = 12$ to the periventricular compartment, while remaining WMH voxels are assigned to the deep white matter compartment. Let $\hat{\mathbf{M}}_{pv}$ and $\hat{\mathbf{M}}_{dw}$ denote the corresponding periventricular and deep WMH masks. The regional involvement ratio is computed as $a_{pv/dw} = |\hat{\mathbf{M}}_{pv}|/(|\hat{\mathbf{M}}_{dw}| + \epsilon)$, where ϵ is a small stability constant set to one voxel to avoid division by zero. Finally, the lesion count a_{cc} is obtained as the number of connected components in $\hat{\mathbf{M}}$. These quantities form an auxiliary target vector $\mathbf{a} = [a_{vol}, a_{pv/dw}, a_{cc}]$.

The auxiliary branch is attached to the fused representation $\mathbf{f}$ and predicts $\hat{\mathbf{a}} = \text{MLP}_{aux}(\mathbf{f})$. We optimize the auxiliary loss as $\mathcal{L}^s_{aux} = \|\hat{\mathbf{a}} - \mathbf{a}\|_1$, where s

Table 1: Quantitative comparison for PVH and DWMH Fazekas grading. AUROC and AUPRC are computed using macro one-vs-rest averaging.

Task	Method	Acc.	F1	AUROC	AUPRC	QWK
PVH	3D ResNet-18	0.686±0.014	0.662±0.017	0.842±0.011	0.742±0.015	0.781±0.013
	3D Swin Transformer	0.721±0.012	0.706±0.015	0.869±0.010	0.789±0.013	0.816±0.011
	2.5D Attention MIL	0.736±0.011	0.724±0.013	0.881±0.009	0.812±0.012	0.837±0.010
	Rule grading	0.704±0.013	0.691±0.016	-	-	0.805±0.012
	RBT-Faz	**0.784±0.033**	**0.786±0.011**	**0.910±0.005**	**0.868±0.011**	**0.886±0.006**
DWMH	3D ResNet-18	0.704±0.013	0.708±0.016	0.861±0.011	0.752±0.015	0.792±0.013
	3D Swin Transformer	0.742±0.011	0.754±0.014	0.895±0.009	0.812±0.012	0.851±0.010
	2.5D Attention MIL	0.765±0.010	0.773±0.012	0.908±0.008	0.829±0.011	0.872±0.009
	Rule grading	0.731±0.012	0.737±0.015	-	-	0.834±0.011
	RBT-Faz	**0.816±0.015**	**0.825±0.007**	**0.937±0.004**	**0.877±0.006**	**0.914±0.005**

denotes the selected Fazekas subscale, either PVH or DWMH. The full objective for target subscale s is:

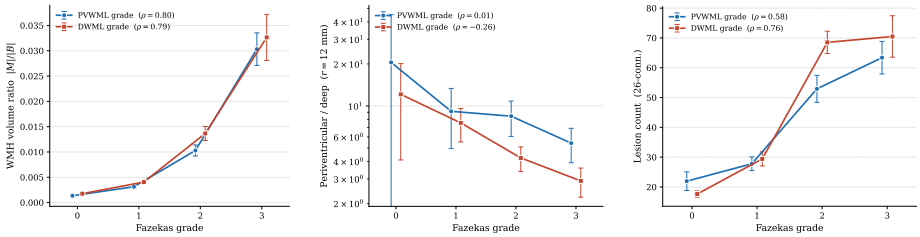
$$\mathcal{L}^s_{total} = \mathcal{L}^s_{cls} + \lambda_{ord}\mathcal{L}^s_{ord} + \lambda_{align}\mathcal{L}^s_{align} + \lambda_{aux}\mathcal{L}^s_{aux}. \tag{5}$$

Here, λ_{ord}, λ_{align}, λ_{aux} are scalar weights balancing the ordinal, alignment, and auxiliary losses relative to $\mathcal{L}^s_{cls}$. The trained model outputs a single ordinal Fazekas subscore $\hat{y}^s \in \{0, 1, 2, 3\}$ for the selected target, either PVH or DWMH.

3 Experiments and Results

Dataset and evaluation protocol. We evaluated the proposed framework on 933 CCNA subjects with 3D FLAIR MRI [14] and radiologist-assigned Fazekas annotations. The original labels were highly imbalanced: PVH grades 0–3 included 38/536/280/79 subjects, while DWMH grades 0–3 included 76/625/181/51 subjects. We therefore used undersampling to construct balanced subsets, yielding train/validation splits of 333/84 (PVH) and 342/85 (DWMH) subjects. To improve volumetric representation learning, the 3D Swin-UNETR encoder was initialized using contrastive pretraining/fine-tuning on additional FLAIR datasets, including a private glioblastoma MRI cohort (GBM, $n = 505$), the Ontario Neurodegenerative Disease Research Initiative (ONDRI, $n = 1468$) [18] and the Alzheimer's Disease Neuroimaging Initiative (ADNI, $n = 4046$) [1]. All images were brain-extracted using a pretrained nnU-Net model [9] and standardized using previously defined methods [6]. For lesion-aware auxiliary supervision, pseudo-WMH and ventricle masks were generated offline using pretrained nnU-Net models. The nnU-Net WMH segmentation model, trained on 527 pooled FLAIR volumes (CCNA, ONDRI, ADNI, MICCAI WMH Challenge [11]), achieved a Dice of 0.723 on held-out data. These masks were used only to compute auxiliary targets, including WMH volume ratio, periventricular-versus-deep WMH ratio with $r = 12$, and connected-component lesion count; they were not provided as inputs to the Fazekas classification model during inference.

Implementation details. The framework was implemented in PyTorch and trained on an NVIDIA RTX 4090 GPU. Separate models were trained for PVH



(a) WMH volume ratio (b) PV/deep WMH ratio (c) Lesion count

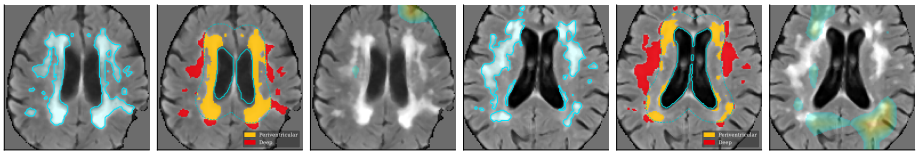
Fig. 2: Clinical grounding of pseudo-mask-derived lesion descriptors across Fazekas grades.

and DWMH using the same architecture, with only the target labels and ClinicalBERT text anchors changed per subscale. Models used AdamW (learning rate 0.0001, batch size 4, 100 epochs). The best checkpoint was selected based on validation QWK. In all experiments, we set $\lambda_{ord} = 0.5$, $\lambda_{align} = 0.1$, and $\lambda_{aux} = 0.1$.

Quantitative comparison. All comparison models were pre-trained, fine-tuned, and evaluated on the same CCNA splits, preprocessing pipeline, and metrics as the proposed method. We compared the proposed framework with 3D ResNet-18 [17], 3D Swin Transformer [21], 2.5D Attention MIL [20], and segmentation-based rule grading [10]. Joo et al. [10] derive a single combined grade from whole-brain WMH volume via ROC/Youden-optimized thresholds. For subscale-specific PVH/DWMH predictions, we applied the same periventricular/deep compartmentalization used for our auxiliary targets ($r = 12$, Sect. 2) and independently refit ROC/Youden-optimal thresholds per compartment on the CCNA training split. As shown in Table 1, the proposed method achieved the strongest performance for both PVH and DWMH grading. The gains in macro-F1 and area under the precision–recall curve (AUPRC) indicate improved performance under class imbalance, while higher quadratic weighted kappa (QWK) demonstrates better ordinal agreement with radiologist-assigned grades.

Component analysis and ordinal error. Component analysis was performed on PVH as a representative task. The Swin-UNETR baseline without FAB, text alignment, or auxiliary supervision achieved 0.711 Macro-F1 and 0.829 QWK. Adding FAB with gated pooling improved performance to 0.748/0.858, and adding text-manifold alignment further raised it to 0.759/0.872. The full RBT-Faz model, with lesion-aware auxiliary supervision, achieved the best Macro-F1/QWK of 0.786/0.886, a relative improvement of 10.5% and 6.9% over the baseline. All incorrect predictions were within one grade of the radiologist-assigned label.

We also evaluated conventional classifiers [5, 2] trained only on pseudo-WMH-derived scalar descriptors (WMH volume, connected-component statistics, periventricular/deep distribution). On the same PVH validation split, the best descriptor-based model, HistGB, achieved 0.672 Macro-F1/0.818 QWK, versus 0.614/0.777



(a) (b) (c) (d) (e) (f)

Fig. 3: PVH and DWMH visualizations. (a)–(c): PVH pseudo-WMH reference mask, periventricular/deep split, and Grad-CAM activation. (d)–(f): corresponding DWMH case.

for GradBoost and 0.581/0.777 for Random Forest. These scalar descriptors underperformed both the Swin-UNETR image baseline and the full RBT-Faz model, suggesting that aggregate handcrafted summaries discard spatial information useful for distinguishing subtle ordinal Fazekas grades.

Clinical grounding and interpretability. We first examined whether pseudo-mask-derived lesion descriptors were clinically consistent with radiologist-assigned severity. As shown in Fig. 2, WMH volume ratio and lesion count increased with Fazekas grade, whereas the periventricular-versus-deep WMH ratio decreased, particularly for DWMH grading, reflecting greater deep white matter involvement in higher-severity cases. These complementary trends support the use of WMH burden, spatial distribution, and lesion count as auxiliary supervision targets. We also visualized model attention using Grad-CAM against pseudo-WMH reference masks and periventricular/deep WMH splits. As shown in Fig. 3, the PVH model activation overlaps with periventricular WMH regions, while the DWMH model attends more strongly to deep WMH involvement, suggesting anatomically meaningful evidence rather than unrelated background regions. Because post-hoc Grad-CAM maps from final classification layers operate on downsampled spatial resolutions, binarized heatmaps naturally yield low absolute voxel-level overlap against crisp segmentation masks. The spatial sensitivity of RBT-Faz is demonstrated by the monotonic increase in overlap with lesion severity (reaching 0.148 for Grade 3 DWMH) and high pathology coverage in finer encoder layers (up to 61%), confirming that the architecture preserves lesion-sensitive spatial evidence.

4 Conclusion

We introduced RBT-Faz, a direct 3D FLAIR MRI framework for ordinal PVH and DWMH Fazekas grading that combines a 3D Swin-UNETR encoder with a register-bottleneck Firewall Attention Block, ClinicalBERT-based text alignment, and lesion-aware auxiliary supervision to improve ordinal consistency and clinical grounding without lesion masks or anatomical priors at inference. RBT-Faz achieved the strongest performance across both subscales on CCNA subjects. Future work includes external validation on an independent held-out test cohort, and broader small vessel disease assessment.

References

1. Aisen, P.S., Petersen, R.C., Donohue, M., Weiner, M.W., Initiative, A.D.N., et al.: Alzheimer's disease neuroimaging initiative 2 clinical core: progress and plans. Alzheimer's & Dementia **11**(7), 734–739 (2015)
2. Breiman, L.: Random forests. Machine learning **45**(1), 5–32 (2001)
3. Brown, R., Low, A., Markus, H.S.: Rate of, and risk factors for, white matter hyperintensity growth: a systematic review and meta-analysis with implications for clinical trial design. Journal of Neurology, Neurosurgery & Psychiatry **92**(12), 1271–1277 (2021)
4. Chen, J., Mikheev, A.V., Yu, H., Gruen, M.D., Rusinek, H., Ge, Y., Initiative, A.D.N., et al.: Bilateral distance partition of periventricular and deep white matter hyperintensities: performance of the method in the aging brain. Academic radiology **28**(12), 1699–1708 (2021)
5. Friedman, J.H.: Greedy function approximation: a gradient boosting machine. Annals of statistics pp. 1189–1232 (2001)
6. Ghazvanchahi, A., Maralani, P.J., Moody, A.R., Khademi, A.: Effect of intensity standardization on deep learning for wml segmentation in multi-centre flair mri. In: Medical Imaging with Deep Learning. pp. 1923–1940. PMLR (2024)
7. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI brainlesion workshop. pp. 272–284. Springer (2021)
8. Huang, K., Altosaar, J., Ranganath, R.: Clinicalbert: Modeling clinical notes and predicting hospital readmission. arXiv preprint arXiv:1904.05342 (2019)
9. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods **18**(2), 203–211 (2021)
10. Joo, L., Shim, W.H., Suh, C.H., Lim, S.J., Heo, H., Kim, W.S., Hong, E., Lee, D., Sung, J., Lim, J.S., et al.: Diagnostic performance of deep learning-based automatic white matter hyperintensity segmentation for classification of the fazekas scale and differentiation of subcortical vascular dementia. PLoS One **17**(9), e0274562 (2022)
11. Kuijf, H.J., Biesbroek, J.M., De Bresser, J., Heinen, R., Andermatt, S., Bento, M., Berseth, M., Belyaev, M., Cardoso, M.J., Casamitjana, A., et al.: Standardized assessment of automatic segmentation of white matter hyperintensities and results of the wmh segmentation challenge. IEEE transactions on medical imaging **38**(11), 2556–2568 (2019)
12. Kuwabara, M., Ikawa, F., Nakazawa, S., Koshino, S., Ishii, D., Kondo, H., Hara, T., Matsuda, S., Maeda, Y., Maeyama, S., et al.: Advantage of grading classification using volumetric artificial intelligence for periventricular hyperintensity and deep subcortical white matter hyperintensity. Scientific Reports **15**(1), 40186 (2025)
13. Markus, H.S., de Leeuw, F.E.: Cerebral small vessel disease: recent advances and future directions. International Journal of Stroke **18**(1), 4–14 (2023)
14. Mohaddes, Z., Das, S., Abou-Haidar, R., Safi-Harab, M., Blader, D., Callegaro, J., Henri-Bellemare, C., Tunteng, J.F., Evans, L., Campbell, T., et al.: National neuroinformatics framework for canadian consortium on neurodegeneration in aging (ccna). Frontiers in neuroinformatics **12**, 85 (2018)
15. Mu, S., Lu, W., Yu, G., Zheng, L., Qiu, J.: Deep learning-based grading of white matter hyperintensities enables identification of potential markers in multi-sequence mri data. Computer Methods and Programs in Biomedicine **243**, 107904 (2024)

16. Phuah, C.L., Chen, Y., Strain, J.F., Yechoor, N., Laurido-Soto, O.J., Ances, B.M., Lee, J.M., Initiative, A.D.N.: Association of data-driven white matter hyperintensity spatial signatures with distinct cerebral small vessel disease etiologies. Neurology **99**(23), e2535–e2547 (2022)
17. Raj, J.R.F., Presskila, X.A., Krishnan, R.S., Sarala, R., Isaac, R.A., Soundiraraj, N.: 3d resnet-18 based deep convolutional network for automated stroke detection in volumetric ct scans. In: 2025 3rd International Conference on Inventive Computing and Informatics (ICICI). pp. 01–08 (2025). https://doi.org/10.1109/ICICI65870.2025.11069487
18. Ramirez, J., Holmes, M.F., Scott, C.J., Ozzoude, M., Adamo, S., Szilagyi, G.M., Goubran, M., Gao, F., Arnott, S.R., Lawrence-Dewar, J.M., et al.: Ontario neurodegenerative disease research initiative (ondri): structural mri methods and outcome measures. Frontiers in neurology **11**, 847 (2020)
19. Rieu, Z., Kim, R.E., Lee, M., Kim, H.W., Kim, D., Yong, J., Kim, J., Lee, M., Lim, H., Kim, J.: A fully automated visual grading system for white matter hyperintensities of t2-fluid attenuated inversion recovery magnetic resonance imaging. Journal of Integrative Neuroscience **22**(3), 57 (2023)
20. Shirkavand, R., Ayromlou, S., Farghadani, S., Tahaei, M.s., Pourakpour, F., Siahlou, B., Khodakarami, Z., Rohban, M.H., Fatehi, M., Rabiee, H.R.: Dementia severity classification under small sample size and weak supervision in thick slice mri. arXiv preprint arXiv:2103.10056 (2021)
21. Tang, Y., Yang, D., Li, W., Roth, H.R., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20730–20740 (2022)
22. Wang, J., Chen, M., Zhang, S., Karaev, N., Schönberger, J., Labatut, P., Bojanowski, P., Novotny, D., Vedaldi, A., Rupprecht, C.: Vggt-ω. arXiv preprint arXiv:2605.15195 (2026)