An Explainable Deep Learning Approach for Frontotemporal Dementia Diagnostic Classification from Structural MRI

Caio dos S. Felipe[1], **Dieine E. B. Schiavon**[1], **Thatiane A. P. Alva**[1], **Carla D. L. Becker**[1]

[1]Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA)
Rua Sarmento Leite, 245 – 90050-170 – Porto Alegre – RS – Brazil

{caio.felipe,dieineb,thatiane,carladiniz}@ufcspa.edu.br

Abstract. *This paper presents an explainable deep learning pipeline for four-class diagnostic classification from T1-weighted structural MRI, discriminating three frontotemporal dementia subtypes from cognitively normal controls. A ConvNeXt-Tiny backbone trained on the Neuroimaging Initiative for Frontotemporal Dementia dataset processes 16 sagittal slices and achieves 84.2% accuracy and macro AUC of 0.942 on a held-out patient cohort. A learned Slice Attention Pooling module exposes per-slice aggregation weights, enabling positional inspection of volume-level decisions. EigenCAM and slice attention profiles yield anatomically consistent explanations for BV, PNFA, and CON, with SV discrimination encoded in the global embedding space.*

1. Introduction

Frontotemporal dementia (FTD) is the third most prevalent form of dementia, preceded only by Alzheimer's disease and Lewy body dementia, with mean age of onset at 56 years, making it a leading cause of early-onset cognitive decline [Antonioni et al. 2023]. No curative or disease-modifying treatment currently exists, and clinical management depends critically on early and accurate subtype identification [Antonioni et al. 2023, National Institute on Aging 2021]. The three principal phenotypes addressed in this work present distinct but overlapping neuroanatomical signatures on structural Magnetic Resonance Imaging (MRI). The behavioral variant (bvFTD) produces atrophy predominantly in the medial prefrontal, orbitofrontal, anterior cingulate, and insular cortices, with high morphological variability across patients and disease stages. The semantic variant (svPPA) is characterized by bilateral anterior temporal atrophy, frequently asymmetric and left-predominant, representing one of the most anatomically consistent patterns in neurodegenerative disease. The nonfluent variant (nfvPPA) affects the left perisylvian anterior cortex, with involvement of frontal and insular structures [Antonioni et al. 2023, Grossman et al. 2023]. Differential diagnosis among these variants requires specialist neuroimaging interpretation and remains subject to inter-observer variability in clinical practice [Whitwell 2019].

Automated classification of FTD subtypes from structural MRI faces two compounding difficulties. First, existing approaches rely predominantly on handcrafted morphometric features extracted through atlas-based pipelines, which require skull stripping, spatial registration, and parcellation as preprocessing steps that may fail or introduce bias when atrophy patterns are atypical [Ma et al. 2024, Metz et al. 2025]. Second, explainability in neuroimaging-based dementia classification has been developed primarily for Alzheimer's disease, where datasets are larger and atrophy patterns more homogeneous;

its systematic application to FTD subtypes, where morphological heterogeneity is the defining challenge, remains limited [Viswan et al. 2024, Martin et al. 2026].

This study investigates whether a diagnostic classification pipeline trained end-to-end on raw sagittal slices, without atlas-based preprocessing, can achieve clinically meaningful discrimination among three FTD subtypes and cognitively normal controls, and whether intrinsic explainability mechanisms can yield anatomically consistent accounts of the resulting decisions.

This paper presents a four-class diagnostic classification pipeline for T1-weighted structural MRI that operates directly on raw sagittal slices, bypassing atlas-based preprocessing and its failure modes on atypical atrophy patterns. A ConvNeXt-Tiny backbone pretrained on ImageNet-1k provides the representational capacity required for transfer learning on a limited labeled dataset. A Slice Attention Pooling (SAP) module aggregates per-slice embeddings through learned scalar weights, exposing the positional contribution of each sagittal position to the volume-level decision, a design choice motivated by auditability rather than predictive performance. EigenCAM is selected as the spatial explainability method [Muhammad and Yeasin 2021] because the depthwise convolutions in ConvNeXt produce sparse gradients, making backpropagation-dependent methods susceptible to anatomically inconsistent activations. Patient-level partitioning before any training or model selection ensures that reported results reflect generalization to unseen patients.

2. Related Works

Among works addressing FTD classification from structural MRI, Ma et al. [Ma et al. 2024] classify bvFTD, svPPA, and nfvPPA using a deep neural network trained on FreeSurfer-derived morphometric features from 277 patients across the Neuroimaging Initiative for Frontotemporal Dementia (NIFD) and Advancing Research and Treatment for Frontotemporal Lobar Degeneration (ALLFTD) datasets, with explainability via integrated gradients over the feature space. The approach operates on handcrafted features rather than raw image data, and explainability is expressed in terms of morphometric variables rather than image-space activations. Hu et al. [Hu et al. 2021] employ a 3D Convolutional Neural Network (CNN) with voxel-based contribution maps to discriminate FTD from Alzheimer's disease using NIFD and Alzheimer's Disease Neuroimaging Initiative (ADNI) data, without performing intra-FTD subtype classification. Metz et al. [Metz et al. 2025] apply partial least squares and machine learning to multivariate neuroimaging features for FTD subtyping across 136 patients from the Frontotemporal Lobar Degeneration Neuroimaging Initiative (FTLDNI) dataset, achieving 89.12% accuracy but without explainability analysis. All three approaches operate on preprocessed morphometric features rather than end-to-end learned representations. Termine et al. [Termine et al. 2022] train a 3D deep learning model directly on NIFD structural MRI data for FTD detection, applying Explainable Artificial Intelligence (XAI) methods to identify regions driving model decisions, but address only binary FTD-versus-control classification without intra-FTD subtype discrimination.

Regarding XAI in dementia neuroimaging, Viswan et al. [Viswan et al. 2024] systematically review XAI methods for Alzheimer's disease classification, establishing the methodological landscape for post-hoc and ante-hoc approaches. Martin et

al. [Martin et al. 2026] benchmark eleven XAI methods for Alzheimer's disease diagnosis, finding gradient-based methods produce the highest inter-method agreement. Singh et al. [Singh et al. 2025] propose an unsupervised XAI framework evaluated on 3253 MRI scans from six cohorts including NIFD, covering cognitively normal controls, Alzheimer's disease, and FTD, with clinical validation by six specialists. None of these works addresses intra-FTD subtype classification specifically. No prior study combines end-to-end learning on raw MRI with intrinsic positional explainability via learned slice attention for FTD diagnostic classification. The present work addresses this gap.

3. Background

3.1. Neuroimaging Patterns in FTD

Structural MRI is routinely used in clinical practice to examine patterns of cerebral atrophy in FTD, and each subtype presents distinctive neuroimaging signatures that make automated image-based classification a well-motivated approach [Whitwell 2019].

In bvFTD, atrophy is observed bilaterally in the medial prefrontal cortex, with specific involvement of the right orbitofrontal cortex, anterior insula, and anterior cingulate cortex, alongside significant reductions in frontal, temporal, and subcortical structures, particularly the striatum [Antonioni et al. 2023, Grossman et al. 2023]. In svPPA, atrophy is characterized by bilateral anterior temporal involvement, frequently asymmetric and left-predominant, progressing posteriorly to affect temporal white matter and inferior frontal cortex, with metabolic changes following an anterior-to-posterior gradient and additional involvement of orbitofrontal and parietal structures [Antonioni et al. 2023, Grossman et al. 2023]. In nfvPPA, structural changes are initially concentrated in the anterior perisylvian cortex of the dominant hemisphere, particularly Broca's area, progressing to involve left frontal regions, the insula, and subcortical structures [Antonioni et al. 2023, Grossman et al. 2023]. Figure 1 illustrates the anatomical distribution of these alterations across the three FTD variants in comparison with healthy controls.

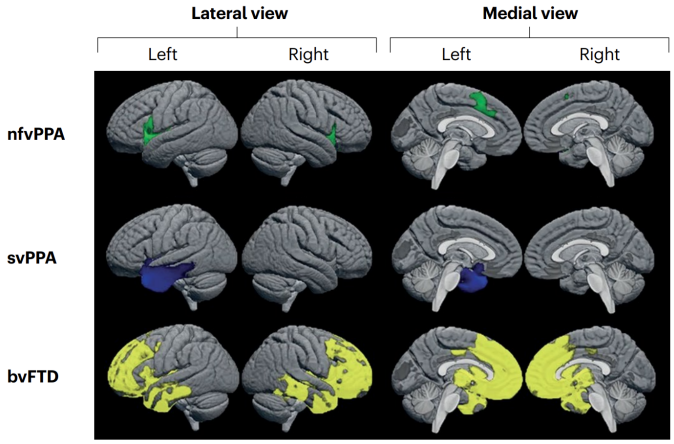


Figure 1. Characteristic neurodegeneration patterns in bvFTD, svPPA, and nfvPPA compared to healthy controls on structural MRI. Adapted from [Grossman et al. 2023].

Detecting these spatially distinct atrophy patterns from raw MRI volumes requires

a backbone capable of capturing both local texture and global structural information across the full sagittal extent of the brain.

3.2. ConvNeXt

ConvNeXt is a family of purely CNN architectures proposed by Liu et al. [Liu et al. 2022] as a systematic modernization of the standard ResNet design, incorporating structural choices derived from Vision Transformers without introducing self-attention mechanisms. The result is a model that retains the simplicity and computational efficiency of convolutional networks while achieving competitive performance on large-scale image classification benchmarks.

Each ConvNeXt block is organized as an inverted bottleneck: a depthwise convolution with a large 7 by 7 kernel is applied first, followed by two pointwise convolutions that expand and then contract the channel dimension by a factor of four. This structure separates spatial mixing, performed by the depthwise convolution, from channel mixing, performed by the pointwise layers, a decomposition also found in Transformer architectures. Layer normalization replaces batch normalization throughout, and a single Gaussian Error Linear Unit (GELU) activation is used per block, reducing the number of nonlinearities relative to earlier convolutional designs. Downsampling between stages is performed by dedicated normalization and strided convolution layers rather than within the residual blocks. Figure 2 illustrates the internal structure of a single ConvNeXt block.

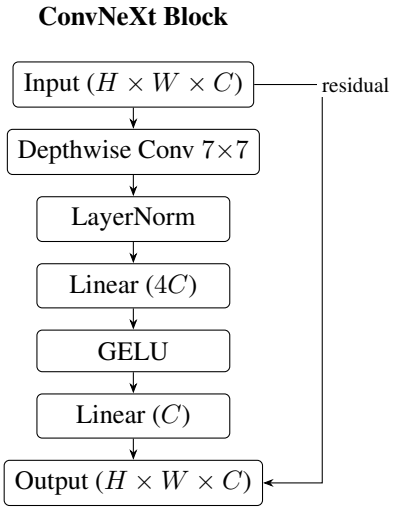


Figure 2. ConvNeXt block structure. A depthwise convolution handles spatial mixing; two pointwise linear layers handle channel mixing with a $4\times$ expansion. Adapted from [Liu et al. 2022].

The architecture is organized into four hierarchical stages with progressively increasing channel widths and decreasing spatial resolution, producing high-level feature maps that capture both local texture and global structural information. ImageNet-pretrained weights are publicly available for all variants, making ConvNeXt well suited for transfer learning in medical imaging settings where labeled data is limited.

4. Materials and Methods

4.1. Dataset

This study uses a subset of the NIFD dataset [NIFD 2010], curated by the University of California San Francisco (UCSF) Memory and Aging Center and publicly available through the Laboratory of Neuro Imaging (LONI) platform, comprising 756 T1-weighted structural MRI series in DICOM format from 257 participants across four diagnostic classes: cognitively normal controls (CON, 340 series, 115 participants), behavioral variant FTD (BV, equivalent to bvFTD, 210 series, 72 participants), semantic variant FTD (SV, equivalent to svPPA, 110 series, 37 participants), and nonfluent variant primary progressive aphasia (PNFA, also referred to as nfvPPA, 96 series, 33 participants). Diagnostic labels are derived from the accompanying clinical data file, linked to each series by patient identifier and examination date. The FTLDNI/NIFD initiative is coordinated by the UCSF Memory and Aging Center, supported by National Institutes of Health (NIH) Grant R01 AG032306, with data distributed by the University of Southern California (USC) Laboratory for Neuro Imaging.

4.2. Data Partitioning and Leakage Prevention

All partitioning is performed at the patient level before any model training or hyperparameter selection. Twenty percent of participants are assigned to a held-out test set, yielding 158 series from 52 participants never used during training or model selection. The remaining 205 participants are used for cross-validation via StratifiedGroupKFold with five folds, using patient identifier as the grouping variable, ensuring that no participant contributes series to both the training and validation subsets of any fold. An overlap assertion is executed programmatically for each fold to confirm zero patient-level contamination. Augmentation is applied exclusively to training subsets.

4.3. Preprocessing

Each DICOM series is converted to a three-dimensional NumPy array and cached to disk. From the resulting volume, 16 sagittal slices are selected at equal intervals via linear spacing, after discarding a margin of 20% from each end of the sagittal axis to avoid predominantly dark border slices. Each selected slice is normalized to the range zero to one by dividing by its maximum value, a choice robust to the consistent intensity scaling of T1-weighted DICOM volumes, then replicated across three channels. Standard ImageNet normalization is subsequently applied, followed by bilinear resampling to 224 by 224 pixels. The resulting input tensor has shape 16 by 3 by 224 by 224 per series.

4.4. Model Architecture

The proposed model processes all 16 slices through a shared ConvNeXt-Tiny backbone initialized from ImageNet-1k weights, producing a 768-dimensional embedding per slice via global average pooling over the final-stage feature maps. The 16 embeddings are aggregated by a Slice Attention Pooling module, which assigns a scalar weight to each slice through a two-layer Multilayer Perceptron (MLP) with Tanh activation, followed by softmax normalization across the 16 positions. The weighted sum produces a single 768-dimensional series-level representation. The classification head applies layer normalization, dropout with probability 0.4, and a linear projection to four output logits. Total parameter count is approximately 28.2 million. Figure 3 summarizes the proposed end-to-end architecture.

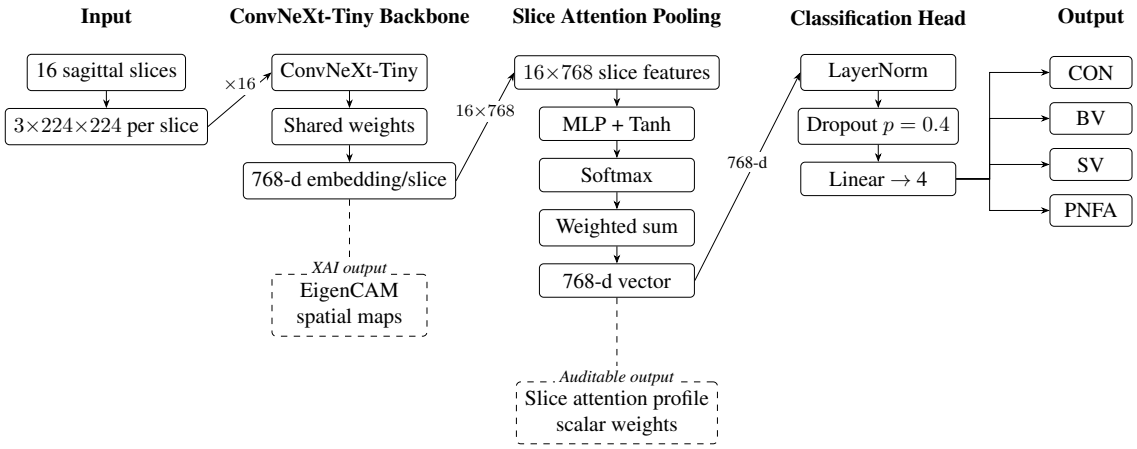


Figure 3. Overview of the proposed ConvNeXt-Tiny and Slice Attention Pooling pipeline for four-class FTD diagnostic classification from structural MRI.

4.5. Training Strategy

Training proceeds in two phases. In Phase 1, spanning the first five epochs, the backbone is frozen at its first two stages and only the attention module and classification head are updated with a learning rate of 3e-4, preventing the randomly initialized head from corrupting pretrained convolutional features. In Phase 2, the backbone is fully unfrozen with differential learning rates: 3e-5 for backbone parameters and 1.5e-4 for the attention module and head. The optimizer is AdamW with weight decay 1e-4, a CosineAnnealingLR scheduler, and gradient clipping with maximum norm 1.0. Training uses a batch size of 4 series. Early stopping monitors validation accuracy with patience 10 over a maximum of 60 epochs. The checkpoint from the fold achieving the highest validation accuracy (Fold 2, val acc $= 0.883$) is used for test set evaluation and XAI analysis. Retraining on the full training split was not performed given the small dataset size, where the marginal increase in training examples per fold does not warrant the associated risk of overfitting to a noisier combined set. The loss function is cross-entropy with class-balanced weights and label smoothing of 0.1. Augmentation applies random horizontal flip, random rotation within 5 degrees, and random affine translation and scaling within conservative bounds, applied consistently across all 16 slices of a series to preserve spatial coherence.

4.6. Evaluation Metrics

Model performance is evaluated using overall accuracy, per-class precision, recall, and F1-score, and macro-averaged AUC via one-versus-rest curves. Cross-validation performance is reported as the mean and standard deviation of per-class F1 and macro F1 across five folds.

4.7. Explainability Methods

Three complementary explainability methods are applied to the best-fold checkpoint: EigenCAM for spatial activation analysis, slice attention analysis for positional contribution, and UMAP projection of the pooled embedding space for population-level structure. EigenCAM is selected as the primary spatial explainability method because depthwise convolutions in ConvNeXt can produce sparse or unstable gradients at the target layer, making gradient-dependent methods susceptible to anatomically inconsistent activations.

EigenCAM decomposes the feature maps of the final convolutional block via principal component analysis, extracting the direction of maximum spatial variance without back-propagation. In medical imaging settings where pathological signal is focal and spatially constrained, this property produces anatomically stable activation maps. Slice attention analysis examines the 16 scalar weights for all test series, averaged by class and by prediction outcome, revealing which sagittal positions the model systematically prioritizes.

A UMAP projection of the 768-dimensional pooled embeddings is computed for all test series to examine the global structure of the learned representation space. Unlike EigenCAM and slice attention analysis, which operate at the level of individual series decisions, UMAP reveals the geometric organization of the full test cohort in the embedding space produced by the SAP module, providing a population-level view of class separability.

5. Results

The model achieves an overall accuracy of 84.2% and a macro AUC of 0.942 on the held-out test set of 158 series across the four classes (CON, BV, SV, PNFA). Given that CON constitutes 53.8% of the test set and is classified with perfect recall, the macro AUC is the primary metric for assessing discriminative capacity across classes. Full per-class metrics are reported in Table 1.

Table 1. Classification report on the held-out test set.

	Precision	Recall	F1-score	AUC	Support
CON	0.810	1.000	0.895	0.948	85
BV	0.947	0.486	0.643	0.880	37
SV	0.909	1.000	0.952	1.000	20
PNFA	0.833	0.625	0.714	0.938	16
Accuracy	N/A	N/A	0.842	N/A	158
Macro avg	0.875	0.778	0.801	0.942	158

The confusion matrix in Figure 4 shows that SV achieves perfect discrimination across all operating thresholds, consistent with the highly characteristic anterior temporal and insular atrophy pattern of semantic dementia. CON produces zero false negatives at the default threshold. BV presents the most significant failure mode: 45.9% of true BV cases are classified as CON, the clinically most consequential error pattern, as it corresponds to misclassifying an FTD subtype as a healthy control. This confusion is consistent with the diffuse and morphologically variable nature of bvFTD atrophy, which can produce structural patterns with limited contrast against normal aging, particularly at early disease stages. PNFA similarly shows high precision but limited recall at the default threshold. The ROC curves in Figure 5 confirm that discriminative capacity for both BV and PNFA substantially exceeds default threshold performance, indicating that operating point adjustment can recover a substantial fraction of missed cases at the cost of reduced precision.

Cross-validation results are reported in Table 2. Folds 1 through 4 produce results consistent with the test set evaluation. Fold 5 shows a substantial drop in accuracy and macro F1. Class imbalance does not account for this result, as Fold 5 contains identical

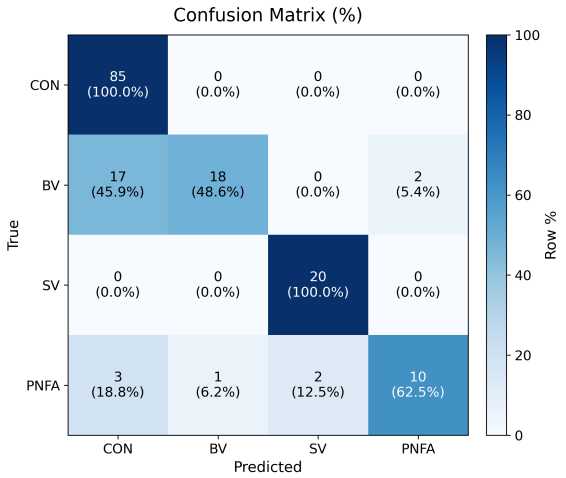


Figure 4. Confusion matrix for the held-out test set (n=158 series).

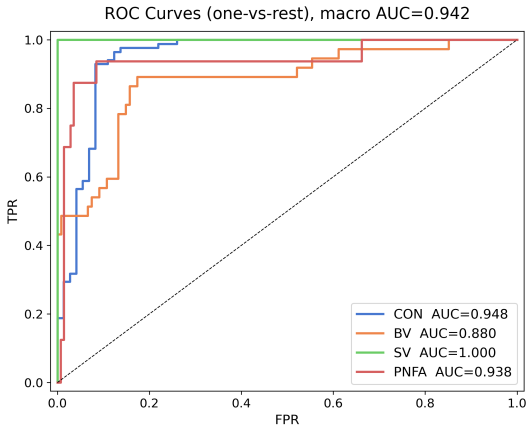


Figure 5. One-versus-rest ROC curves for the four diagnostic classes on the held-out test set.

class counts and the lowest imbalance ratio among all folds. One plausible hypothesis is inter-patient variability, with patients in that partition presenting atrophy patterns more atypical or phenotypically overlapping than those in the remaining folds; however, the dataset size precludes definitive characterization of this fold's composition. This worst-case performance quantifies the sensitivity to patient population composition that should be expected in practice.

Table 2. Five-fold cross-validation results. Fold 2 was selected as the best model checkpoint.

Fold	Val Acc	CON F1	BV F1	SV F1	PNFA F1	Macro F1
1	0.850	0.921	0.730	0.905	0.743	0.807
2	0.883	0.924	0.762	1.000	0.733	0.855
3	0.842	0.919	0.667	0.973	0.647	0.802
4	0.849	0.922	0.563	0.974	0.718	0.794
5	0.723	0.901	0.457	0.683	0.323	0.591
Mean ±SD	0.829 ±0.055	0.917 ±0.010	0.636 ±0.121	0.907 ±0.119	0.633 ±0.170	0.770 ±0.096

Table 3. Slice Attention Pooling (SAP) versus Global Average Pooling (GAP) ablation on the held-out test set. Both models use a ConvNeXt-Tiny backbone, identical training strategy, and the same patient-level split.

Model	Acc	Macro F1	Macro AUC	BV Rec	BV F1	PNFA F1
ConvNeXt-Tiny + SAP	0.842	0.801	0.942	0.486	0.643	0.714
ConvNeXt-Tiny + GAP	0.848	0.807	0.961	0.541	0.667	0.647

Table 3 reports the Slice Attention Pooling versus Global Average Pooling ablation. Replacing Slice Attention Pooling with uniform Global Average Pooling over slice embeddings produced slightly higher accuracy, macro F1, and macro AUC on the held-out test set. SAP therefore did not confer a predictive performance advantage in this dataset. SAP is retained as an auditability-oriented aggregation mechanism: the learned slice-level weights expose the positional contribution of each sagittal position to the volume-level representation, enabling the per-class attention profiles shown in Figure 6. These weights are not interpreted as causal explanations of model behavior.

SV cases concentrate attention on anterior slice positions, consistent with temporal pole and anterior temporal lobe atrophy. BV cases show a broader and more variable distribution, reflecting the diffuse nature of behavioral variant atrophy. PNFA attention concentrates in anterior-to-mid sagittal positions, consistent with the frontal opercular and perisylvian involvement characteristic of nfvPPA. CON attention weights are distributed more uniformly, consistent with the absence of focal structural changes in healthy subjects.

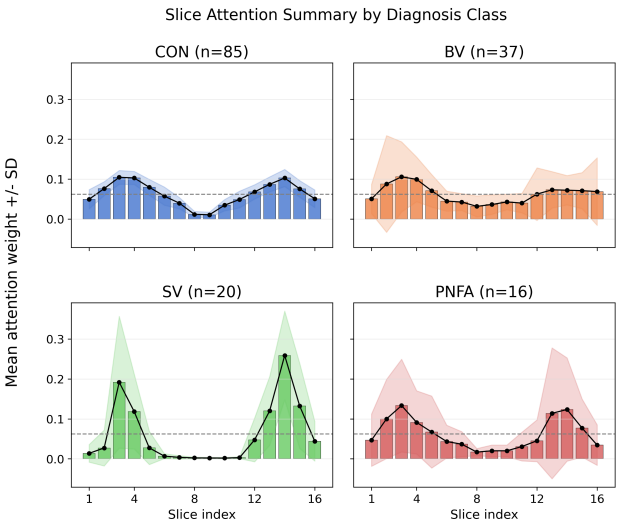


Figure 6. Mean slice attention weights per diagnostic class across the 16 sagittal positions. The dashed horizontal line represents the uniform baseline (1/16). Bars above the baseline indicate positions the model considers more discriminative for that class.

Figure 7 shows that SV embeddings form a compact, fully separated cluster with no overlap with any other class, consistent with the perfect AUC reported in Table 1. CON and BV embeddings share a common manifold with substantial overlap, providing a geometric account of the 45.9% misclassification rate between BV and CON at the

default threshold. PNFA occupies a distinct region adjacent to BV, consistent with its shared frontal involvement.

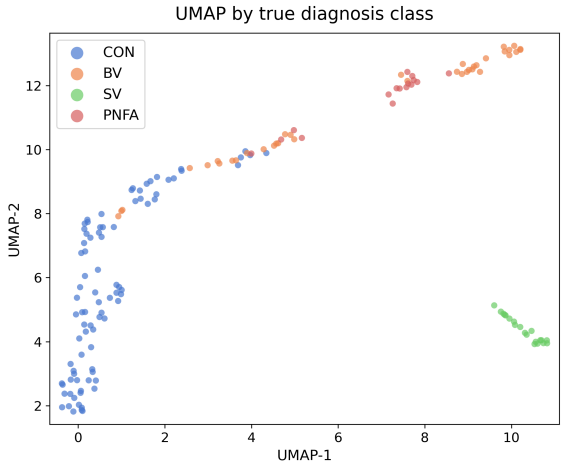


Figure 7. UMAP projection of the 768-dimensional pooled embeddings for all test series (n=158), colored by true diagnostic class. Each point represents one MRI series.

EigenCAM activation maps, shown in Figure 8, reveal class-specific spatial patterns consistent with known FTD neuroanatomy. CON maps show the most diffuse pattern, with broad central-parietal activation and no anterior focal signal. BV activations shift anteriorly and inferiorly relative to CON, concentrating in the frontal and orbitofrontal cortex, consistent with prefrontal degeneration in bvFTD. SV maps do not produce a visually distinct focal pattern relative to CON. This is consistent with the UMAP structure in Figure 7, where SV discrimination is encoded in the global pooled embedding rather than in spatially localized activations within individual sagittal slices, a regime in which EigenCAM, operating on per-slice feature maps, is not the appropriate instrument. PNFA activations are the most superiorly and anteriorly concentrated of all four classes, corresponding to the superior frontal and perisylvian cortex, consistent with the frontal opercular and Broca's area involvement of nfvPPA.

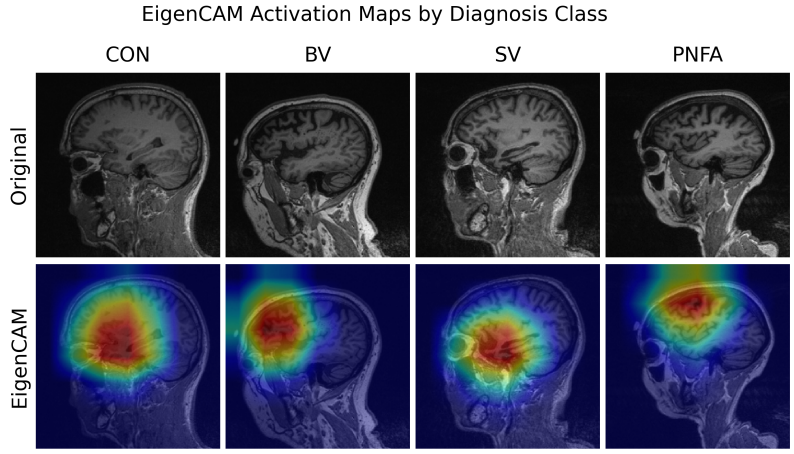


Figure 8. EigenCAM activation maps for representative correctly classified cases of each diagnostic class. Warmer colors indicate higher activation.

6. Conclusion

This paper presented an atlas-free explainable deep learning pipeline for four-class FTD diagnostic classification from structural MRI. EigenCAM spatial analysis produces anatomically consistent explanations for BV, PNFA, and CON, while UMAP of the pooled embeddings accounts for SV discrimination, which is encoded in the global representation space and does not produce spatially localized per-slice activations. Slice attention profiles complement both methods by exposing the positional contribution of each sagittal position to the volume-level decision, and these weights are not interpreted as causal explanations of model behavior.

These results carry direct implications for differential diagnosis in clinical practice. The confusion between BV and CON identified in Figure 4 and the shared embedding manifold shown in Figure 7 indicate that the model's greatest clinical risk lies in the behavioral variant, where diffuse and early-stage atrophy can resemble normal aging on structural MRI. In a diagnostic support role, this suggests the model would be best positioned as a triage or second-reader tool that flags cases for specialist review, rather than as a standalone classifier, particularly for patients presenting with the heterogeneous frontal profile characteristic of bvFTD. By contrast, the near-perfect discrimination of SV (Table 1, Figure 7) suggests that automated support could meaningfully reduce specialist workload for cases with the anatomically consistent anterior temporal pattern typical of semantic variant PPA, where model confidence and clinical presentation are most likely to agree.

Several limitations constrain the generalizability of these findings. The clinically most significant is the conservative BV recall at the default operating threshold, which reflects genuine morphological heterogeneity in bvFTD and should be addressed in future work through larger and more phenotypically diverse training cohorts. The model was also trained and evaluated on a single dataset (NIFD), and no empirical comparison against alternative published methods for FTD subtype classification was performed. The dataset size, particularly for the SV (n=20) and PNFA (n=16) classes in the test set, limits the statistical power of per-class estimates. Cross-validation Fold 5 showed a substantial performance drop not explained by class imbalance, suggesting sensitivity to patient population composition that the dataset size does not allow to be fully characterized. Finally, slice attention weights and EigenCAM activations are presented as auditability tools rather than validated causal explanations of model behavior. Expert clinical review of these explanations was not performed and remains necessary before any diagnostic application.

External validation on multicenter data remains the necessary condition for any claim about generalizability and constitutes the primary direction for future work.

References

Antonioni, A., Raho, E. M., Lopriore, P., Pace, A., Spina, E., et al. (2023). Frontotemporal dementia, where do we stand? a narrative review. *International Journal of Molecular Sciences*, 24(14):11732.

Grossman, M., Seeley, W. W., Boxer, A. L., et al. (2023). Frontotemporal lobar degeneration. *Nature Reviews Disease Primers*, 9(40).

Hu, J., Qing, Z., Liu, R., Zhang, X., Lv, P., Wang, M., et al. (2021). Deep learning-based classification and voxel-based visualization of frontotemporal dementia and alzheimer's disease. *Frontiers in Neuroscience*, 14:626154.

Liu, Z., Mao, H., Wu, C.-Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A ConvNet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11976–11986.

Ma, D., Stocks, J., Rosen, H., Kantarci, K., Lockhart, S. N., Bateman, J. R., et al. (2024). Differential diagnosis of frontotemporal dementia subtypes with explainable deep learning on structural MRI. *Frontiers in Neuroscience*, 18:1331677.

Martin, S. A., Zhao, A., Qu, J., Imms, P., Irimia, A., Barkhof, F., and Cole, J. H. (2026). Investigating the utility of explainable artificial intelligence for neuroimaging-based dementia diagnosis and prognosis. *Human Brain Mapping*, 47(2):e70456.

Metz, A., Zeighami, Y., Ducharme, S., Villeneuve, S., and Dadar, M. (2025). Frontotemporal dementia subtyping using machine learning, multivariate statistics and neuroimaging. *Brain Communications*, 7(1):fcaf065.

Muhammad, M. B. and Yeasin, M. (2021). Eigen-CAM: Visual explanations for deep convolutional neural networks. *SN Computer Science*, 2(1):47.

National Institute on Aging (2021). Frontotemporal disorders: Information for patients, families, and caregivers.

NIFD (2010). NIFD: Neuroimaging in frontotemporal dementia database. Laboratory of Neuro Imaging (LONI), USC. `https://ida.loni.usc.edu` (accessed on 08 Jun 2026).

Singh, D., Brima, Y., Levin, F., Becker, M., Hiller, B., Hermann, A., et al. (2025). An unsupervised XAI framework for dementia detection with context enrichment. *Scientific Reports*, 15(1):39554.

Termine, A., Fabrizio, C., Caltagirone, C., and Petrosini, L. (2022). A reproducible deep-learning-based computer-aided diagnosis tool for frontotemporal dementia using MONAI and clinica frameworks. *Life*, 12(7):947.

Viswan, V., Shaffi, N., Mahmud, M., Subramanian, K., and Hajamohideen, F. (2024). Explainable artificial intelligence in alzheimer's disease classification: A systematic review. *Cognitive Computation*, 16(1):1–44.

Whitwell, J. L. (2019). Ftd spectrum: Neuroimaging across the FTD spectrum. In Becker, J. T. and Cohen, A. D., editors, *Brain Imaging*, volume 165 of *Progress in Molecular Biology and Translational Science*, pages 187–223. Elsevier.