Spherical Clinical Embedding with Adaptive Multi-Expert Fusion for AAA Rupture Classification

Merjulah Roby[1⋆], Abu Noman Md Sakib[1⋆], Zijie Zhang[1], Satish C. Muluk[2], Mark K. Eskandari[3], Vikram S. Kashyap[3], and Ender A. Finol[1]

[1] The University of Texas at San Antonio, San Antonio, TX, USA
[2] Allegheny General Hospital, Pittsburgh, PA, USA
[3] Northwestern University, Chicago, IL, USA
merjulah.roby@utsa.edu
⋆These authors contributed equally to this work.

Abstract. Abdominal aortic aneurysm (AAA) rupture risk prediction remains a complex and clinically important task, largely due to substantial variability in morphology, texture characteristics, and the frequent presence of incomplete clinical information. Many existing approaches rely on single-modality representations or static fusion strategies, which can limit robustness when data are missing or partially observed. To address these limitations, we propose an adaptive multi-expert multimodal fusion framework for AAA rupture classification. The proposed approach includes geometric measurements, region-based radiomics texture features, deep semantic representations extracted from mask-guided CTA slices, and structured clinical variables within an uncertainty-guided routing architecture. A missingness-aware spherical clinical embedding module jointly models clinical feature values and their corresponding missingness patterns, enabling stable representation learning under incomplete observations. In addition, a unified cross-modal expert is introduced to capture higher-order interactions between imaging and clinical representations. An adaptive routing mechanism further refines the fusion process by dynamically weighing expert contributions according to confidence-driven descriptors, allowing case-specific integration under heterogeneous structural and clinical conditions. We evaluated XMoE on a private AAA dataset and the publicly available OASIS dataset. XMoE achieves 91.25% accuracy on AAA, outperforming state-of-the-art multimodal fusion approaches, and 88.08% accuracy on OASIS. In addition, it maintains 86.25% accuracy on AAA with 30% simulated missing modality conditions.

Keywords: Abdominal Aortic Aneurysm · Rupture Risk Prediction · Deep Learning · CTA Imaging · Multi-model Fusion · Radiomics.

1 Introduction

Abdominal aortic aneurysm (AAA) is a progressive vascular disease with high morbidity and mortality rates due to the potential for catastrophic rupture [15].

Although clinical decisions are heavily based on maximum diameter, rupture risk is fundamentally driven by complex biomechanical and tissue-level processes [16, 13]. The risk of rupture is affected by a variety of factors, such as wall stress distribution, thrombus heterogeneity, geometric asymmetry, and localized degeneration, which require the development of computational models. Finite element modeling has shown that peak wall stress can potentially outperform diameter in rupture risk prediction [2]. Although such models are physiologically well-motivated, they are highly dependent on segmentation, meshing, and material modeling assumptions, making them less scalable for practical clinical use. In contrast, our framework does not require biomechanical simulation or mesh generation. Automatically generated segmentation masks are used only as a soft attention prior for image modulation and radiomics feature extraction. This provides a lightweight localization mechanism without relying on physics-based modeling. More recently, the field of radiomics has allowed high-throughput extraction of quantitative imaging features from CT angiography, uncovering correlations between texture heterogeneity and aneurysm instability [8]. However, traditional radiomics analysis pipelines have been plagued by issues of low reproducibility and high variability with acquisition.

In recent years, deep learning has significantly advanced medical image analysis. Convolutional neural networks (CNNs) have shown state-of-the-art performance in a variety of radiological tasks, including detection, segmentation, and prediction in oncology [5, 23]. In particular, foundation models and large-scale pretraining have shown enhanced robustness and generalization in medical imaging tasks [1]. However, AAA-related deep learning research has focused primarily on segmentation or diameter prediction, not rupture classification, and is often based on single-modality image representations. AAA rupture or symptoms are manifested by complex interactions among multi-modal imaging features, geometric representations, and clinical data. Although transformer architectures have shown improved performance in medical image segmentation tasks [19], segmentation alone does not fully represent the range of risk factors for rupture. Because clinical records are often incomplete, standard multimodal classification methods often degrade. Projecting missing tabular features into a unit-norm spherical embedding provides stability against magnitude shifts [12, 22]. As such, multi-modal fusion models that combine imaging data with complementary clinical data have shown improved predictive performance in radiology and computational pathology [6]. Mixture-of-Experts (MoE) models provide adaptive routing mechanisms that dynamically weight expert contributions, allowing for a robust combination of heterogeneous feature spaces [17]. By selectively activating specialized experts, such models improve representational flexibility while maintaining computational efficiency. Recent mixture-of-experts methods target missing or arbitrary modality combinations, including Flex-MoE [21] and FuseMoE [4], as well as a multi-task formulation that jointly models patients, modalities, and tasks [20]. These methods handle modality dropout through learned gating, whereas our framework adds a missingness-aware spherical clinical embedding and confidence-driven routing derived from expert logits. Another

important challenge in CT-based aneurysm analysis is the need to maintain anatomical context while focusing on pathological aspects. Hard cropping of aneurysmal areas may lose important global structural information, while soft mask-guided modulation allows for focus on the aneurysmal walls while maintaining the surrounding vascular information. Attention-based multi-instance learning paradigms have proven to be highly effective in histopathological tasks with patch-based representations of variable length [7].

To overcome these challenges, we present an uncertainty-guided multi-expert framework for AAA rupture risk classification. **The contributions of this work are:** 1) We propose an adaptive multi-modal fusion framework for AAA rupture classification that integrates geometric parameters together with region-specific radiomics texture features, deep semantic representations derived from mask-guided CT slices, and structured clinical variables for comprehensive rupture risk assessment. 2) We introduce a missingness-aware spherical clinical embedding module that jointly models clinical variables and their corresponding missingness mask, performs data-driven feature reconstruction, and constrains latent representations to a unit-norm space to ensure stable learning under incomplete clinical data. 3) We develop an uncertainty-guided adaptive routing mechanism that computes confidence-based routing weights from expert logits and applies temperature-scaled softmax fusion, enabling dynamic expert weighting and improved robustness under heterogeneous geometric, semantic, and clinical conditions. Our implementation is available at `https://github.com/anmspro/xmoe-miccai`.

2 Methodology

Figure 1 shows the proposed XMoE framework, which includes multiple expert modules for processing different modalities, with dynamic routing to combine expert outputs.

Structured Descriptor Expert. AAA rupture likely depends on geometric deformation, wall texture instability, and patient-specific clinical characteristics. To model these complementary factors, we construct a structured descriptor that combines geometric and radiomics features extracted from images. Let $\mathcal{V}$ denote the reconstructed aneurysm volume and $\Gamma(\mathcal{V})$ its surface. Since maximum diameter alone cannot capture three-dimensional irregularities or surface variability related to wall stress, we define a geometric vector $\mathbf{g}$ comprising volume $V(\mathcal{V})$, surface area $A(\Gamma(\mathcal{V}))$, maximum extent $D_{\max}(\Gamma(\mathcal{V}))$, and sphericity $\Phi(\mathcal{V})$. The heterogeneity of the wall and thrombus texture is measured by applying L-level gray quantization $Q_L(\mathbf{X}_k)$ to each slice and computing first-order and higher-order texture features from GLCM, GLRLM, and GLSZM within anatomical subregions $r \in \{\text{lumen}, \text{outer wall}\}$. The subregion embeddings $\mathbf{h}_{k,r}$ are then concatenated to obtain the image-level texture representation $\mathbf{z}_k$. Clinical variables are standardized for consistency for all patients. Finally, $\mathbf{g}$ and $\mathbf{z}_k$

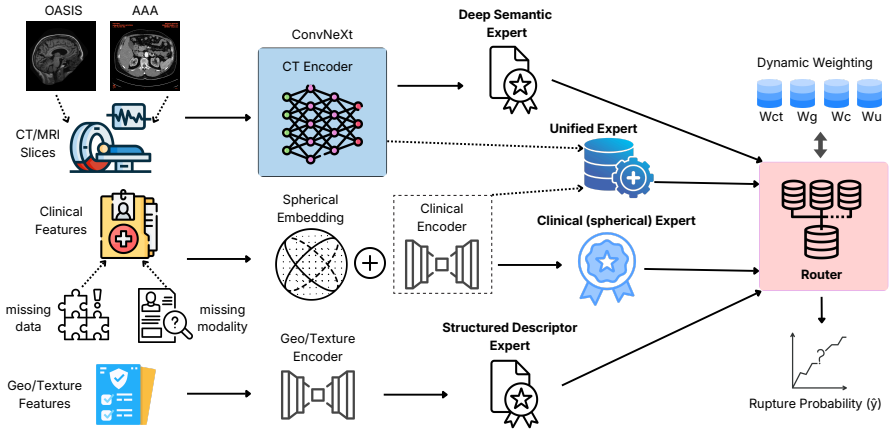


Fig. 1: The overview of the proposed XMoE framework. The diagram shows the integration of multimodal data (CT/MRI, clinical, and geometric features), dynamic weighting by the router, and the final prediction of rupture probability. Each expert module processes a specific modality, with missing data handled by the spherical embedding and imputation module.

are concatenated into a structured representation $\mathbf{s}_k$, which is processed by a dedicated multilayer perceptron to produce the structured expert logit.

Deep Semantic Expert. To balance preservation of anatomical context with emphasis on pathological regions, each slice $\mathbf{X}_k$ is adaptively modulated using its corresponding mask $\mathbf{M}_k$, yielding $\tilde{\mathbf{X}}_k = \mathbf{X}_k \odot (\alpha + (1-\alpha)\mathbf{M}_k)$, where $\alpha \in (0,1)$ controls background retention. This formulation enhances the regions of interest while retaining sufficient contextual information. Semantic features are extracted using a ConvNeXt-Tiny backbone $\Psi(\cdot)$, and a slice-level embedding is obtained via masked feature aggregation and global pooling, producing $\mathbf{u}_k$. The embedding $\mathbf{u}_k$ is then passed through a prediction head to generate a deep semantic logit ℓ_d, capturing high-level contextual and morphological patterns beyond handcrafted descriptors.

Clinical (spherical) expert. Clinical variables are represented as a feature vector $\mathbf{t} \in \mathbb{R}^d$ together with a binary missingness mask $\mathbf{m} \in \{0,1\}^d$, where $m_j = 1$ indicates observed entries and $m_j = 0$ denotes missing values. Rather than applying naive imputation, we employ a spherical tabular embedding module that jointly processes $(\mathbf{t}, \mathbf{m})$ to produce both an imputed feature estimate $\hat{\mathbf{t}}$ and a normalized latent embedding $\mathbf{z}$. The effective input is constructed by preserving observed entries and replacing missing ones using $\hat{\mathbf{t}}$, ensuring that reliable information is retained while missing values are reconstructed in a data-driven manner. The embedding $\mathbf{z}$ is constrained to a unit-norm space, which stabilizes

representation learning under varying sparsity by emphasizing directional consistency rather than feature magnitude. The reconstructed feature vector is subsequently passed through an MLP to generate a clinical logit, while an auxiliary imputation objective is selectively applied to the observed dimensions ($m_j = 1$) to encourage consistent feature reconstruction. By explicitly incorporating $\mathbf{m}$ into both reconstruction and representation learning, this expert captures missingness as a structured signal rather than noise to improve robustness.

Unified expert. To model cross-modal dependencies, we introduce a unified expert that jointly processes imaging and clinical representations. Let $\mathbf{f} \in \mathbb{R}^{d_f}$ denote the CT feature embedding extracted from the backbone network; the unified expert constructs a joint representation by concatenating $\mathbf{f}$ with the clinical vector $\mathbf{t}$ and its mask $\mathbf{m}$, forming a composite input that is passed through a multi-layer perceptron to produce a unified logit. This design enables the model to capture higher-order interactions between modalities, where the interpretation of imaging features $\mathbf{f}$ is conditioned on patient-specific clinical context $\mathbf{t}$. The inclusion of $\mathbf{m}$ allows the network to distinguish between observed and imputed entries. It prevents unreliable clinical signals from dominating the prediction when the missingness is high. Compared to independent experts, this joint formulation allows the model to learn context-aware feature dependencies, while the mixture-of-experts framework dynamically adjusts its contribution based on the reliability of $\mathbf{f}$ and $(\mathbf{t}, \mathbf{m})$. The resulting outcome contains improved robustness under both partial and complete modality absence.

Adaptive Routing and Fusion. Static fusion assumes equal reliability among experts. However, in practice, rupture evidence may manifest differently among patients, with some cases dominated by geometric deformation and others driven by semantic or clinical patterns. To adaptively weight expert contributions, we construct routing descriptors based on expert confidences. Let $\sigma(\cdot)$ denote the sigmoid function. Expert logits ℓ_{ct}, ℓ_{geo}, ℓ_{clin}, and ℓ_{uni} are derived from slice-level image features and subject-level structured and clinical representations. The corresponding expert confidences are defined as $p_i = \sigma(\ell_i)$. The routing vector is constructed as $\mathbf{r}_k = [|\sigma(\ell_{\text{ct}}) - 0.5|, \sigma(\ell_{\text{ct}}), \sigma(\ell_{\text{geo}}), \sigma(\ell_{\text{clin}}), \sigma(\ell_{\text{uni}}), |\sigma(\bar{\ell}) - 0.5|]$, where $\bar{\ell} = \frac{1}{4}\sum_i \ell_i$ denotes the average logit for the four experts (ct, geo, clin, uni), capturing global expert agreement. Routing weights are computed using a temperature-scaled softmax $\mathbf{w}_k = \text{softmax}(f_r(\mathbf{r}_k)/\tau)$, where τ controls distribution sharpness. The final fused logit for each subject is obtained as a weighted combination $f_k = \sum_{i \in \{ct,geo,clin,uni\}} w_{k,i}\ell_i$. This adaptive mechanism enables dynamic expert selection, allowing the model to emphasize the most reliable predictive pathway under varying structural, semantic, and clinical conditions. The adaptively fused prediction is optimized using binary cross-entropy with logits, $\mathcal{L}fusion = \text{BCEWithLogits}(f_k, y)$. This supervision helps the routing mechanism dynamically weight the most trustworthy expert route. Although image features are extracted at the slice level, they are combined using mean pooling to produce a single subject-level prediction. The overall training ob-

jective combines branch-level supervision and auxiliary regularization terms, $\mathcal{L} = \mathcal{L}fusion + \mathcal{L}uni + \mathcal{L}c + \lambda_s \mathcal{L}sphere + \lambda_c \mathcal{L}_{contrastive}$, enabling end-to-end optimization of expert predictions and adaptive routing weights. The binary cross-entropy loss functions for the unified and clinical experts at the branch level are denoted by L_{uni} and L_c, respectively.

3 Experiments and Results

3.1 Datasets

We retrospectively collected CTA scans from 95 patients with ruptured and unruptured AAAs following Institutional Review Board (IRB) approval. AAA segmentation was performed using a known automated segmentation tool [14]. The dataset includes 4,550 unruptured and 4,627 ruptured AAA images, collected with a waiver of informed consent. The evaluation was performed using 5-fold stratified cross-validation at the patient level. All slices from the same patient were kept within the same fold to prevent data leakage. To assess the generalizability of our approach beyond the private AAA dataset, we additionally evaluated the publicly available OASIS-1 dataset [11]. A total of 436 subjects were included, with binary labels derived from the Clinical Dementia Rating (CDR) score (0 vs. > 0). Available modalities include structural MRI and tabular clinical variables such as age, sex, education, MMSE, and brain volume measurements. This dataset differs significantly from AAA in both imaging modality and clinical context, providing a robust testbed for evaluating the general applicability of the proposed framework. We do not claim clinical transferability between dementia and AAA rupture. OASIS is included solely to test whether the adaptive routing and spherical embedding remain effective under a different imaging modality and clinical context with incomplete data.

3.2 Performance Comparison

We evaluated performance using Accuracy, Precision, F1-score, Matthews Correlation Coefficient (MCC), Sensitivity, and Specificity. We compared our method with four recent multimodal learning approaches: MMD [3], MHDGCN [10], MHGSA [9], and TAMME [18]. These methods represent diverse strategies for multimodal fusion, including contrastive learning, graph-based modeling, attention mechanisms, and transformer-based architectures. All baselines were implemented under the same data splits and preprocessing pipeline for fair comparison.

Tables 1 and 2 show the comparison of performance metrics. On the AAA dataset, the proposed XMoE model achieves the best performance in all metrics. In particular, the model achieves the highest accuracy (0.9125) and MCC (0.8337) with significant improvements in specificity. Furthermore, the low variance observed across all folds demonstrates the framework's robustness to patient variations. On the OASIS dataset, XMoE also achieves the highest accuracy and specificity, while maintaining competitive sensitivity. Despite the differences in modality and clinical differences, the model has consistent performance.

Table 1: Performance comparison on the AAA dataset.

Methods	***Acc***	***Pre***	***F1***	***MCC***	***Sen***	***Spe***
MMD [3]	0.8750	0.9444	0.8606	0.7659	0.8000	0.9500
MHDGCN [10]	0.6625	0.7314	0.4722	0.3786	0.3750	0.9500
MHGSA [9]	0.7875	0.8464	0.7373	0.5877	0.6750	0.9000
TAMME [18]	0.8000	0.8189	0.8019	0.6106	0.8000	0.8000
XMoE (Ours)	**0.9125**	**0.9714**	**0.9048**	**0.8337**	**0.8500**	**0.9750**

Table 2: Performance comparison on the OASIS dataset.

Methods	***Acc***	***Pre***	***F1***	***MCC***	***Sen***	***Spe***
MMD [3]	0.8578	0.6285	0.7526	0.6854	**0.9400**	0.8334
MHDGCN [10]	0.8211	0.5779	0.7015	0.6147	0.9000	0.7975
MHGSA [9]	0.8485	0.6504	0.7469	0.6769	0.9100	0.8300
TAMME [18]	0.8234	0.5764	0.7058	0.6257	0.9200	0.7946
XMoE (Ours)	**0.8808**	**0.7437**	**0.7932**	**0.7408**	0.9000	**0.8751**

3.3 Ablation Study

We performed ablations to evaluate the contribution of individual components of the proposed model. Tables 3 and 4 show that removing the spherical clinical expert leads to a substantial drop in performance on both datasets, confirming its importance in handling missing feature-level data. Without this component, the model becomes significantly more sensitive to incomplete clinical inputs. Similarly, removing the unified expert results in further performance loss, highlighting the importance of modeling cross-modal interactions. This indicates that while modality-specific experts capture independent signals, the unified expert is essential to learn contextual dependencies between imaging and clinical features. Overall, the full XMoE model consistently achieves the best performance, demonstrating that both the spherical embedding mechanism and the unified cross-modal expert are critical to the effectiveness of proposed framework.

3.4 Robustness to Missing Data and Missing Modality

To evaluate the robustness of our framework in realistic clinical scenarios, we simulated missing data and modality conditions by progressively increasing the missingness rate from 5% to 30% in the AAA and OASIS datasets. Missingness was simulated at the patient level under a missing-completely-at-random (MCAR) protocol. Each non-imaging modality was independently removed with probability 5-30% while CT was retained, and baselines received the same zero-imputed inputs. As illustrated in Figure 2, XMoE demonstrates significantly more stable behavior compared to the baseline approaches. In the missing data setting, XMoE maintains high accuracy even at elevated missing rates. It clearly

Table 3: AAA ablation results.

Configuration	*Acc*	*Pre*	*F1*	*MCC*	*Sen*	*Spe*
W/O Spherical Expert	0.7839	0.8231	0.8066	0.5882	0.8250	0.7429
W/O Unified Expert	0.7446	0.7817	0.7668	0.5009	0.7750	0.7143
Full Model	**0.9125**	**0.9714**	**0.9048**	**0.8337**	**0.8500**	**0.9750**

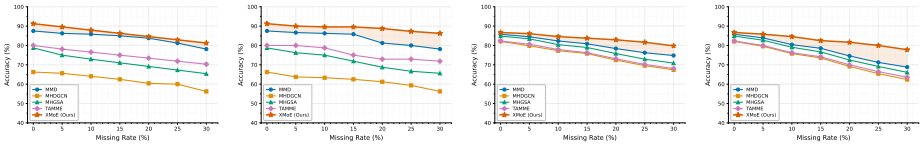
Table 4: OASIS ablation results.

Configuration	*Acc*	*Pre*	*F1*	*MCC*	*Sen*	*Spe*
W/O Spherical Expert	0.7820	0.5060	0.6108	0.4993	0.7900	0.7796
W/O Unified Expert	0.7958	0.5533	0.6592	0.5621	0.8500	0.7795
Full Model	**0.8808**	**0.7437**	**0.7932**	**0.7408**	**0.9000**	**0.8751**

shows the effective handling and imputation of incomplete clinical features. In the missing modality setting, baseline methods exhibit sharp performance degradation due to their reliance on fixed fusion strategies. In contrast, XMoE maintains a superior accuracy of 0.8625 under the exact same 30% missing modality condition. This resilience is primarily driven by the dynamic routing mechanism, which adaptively reweighs modality-specific experts based on feature availability, alongside the spherical clinical expert that stabilizes feature representations under partial observations.

4 Conclusion

We proposed an adaptive multi-expert multimodal learning framework for AAA rupture classification that combines geometrical features, deep semantic features extracted from mask-guided CTA images, and missingness-aware spherical clinical embeddings in a single routing framework. Our proposed framework utilizes uncertainty-driven adaptive fusion to learn dynamic weights for each expert based on case-level confidence, allowing for reliable prediction in diverse structural and clinical settings. The spherical clinical component promotes stable learning from incomplete observations by modeling joint feature reconstruction and representation learning, and the unified expert component captures intermodal relationships between imaging and tabular modalities. By extensive evaluation on our private AAA dataset and the publicly available OASIS dataset, we showed clear gains over state-of-the-art multimodal fusion approaches in accuracy, MCC, and robustness to missing data and missing modality scenarios. The framework with the available modalities during inference can be used in a clinical setting. A limitation of this study is the relatively small single-cohort dataset. Also using clinically structured (informative) missingness would have been more realistic. Future work will explore validation on larger multi-center datasets and multimodal modeling.



(a) AAA: Data (b) AAA: Modality (c) OASIS: Data (d) OASIS: Modality

Fig. 2: Performance comparison of XMoE and baseline models under varying rates (5% to 30%) of simulated data incompleteness. For the AAA and OASIS datasets, XMoE exhibits superior robustness to missing intra-modality data and entire missing modalities compared to fixed-fusion baselines.

Acknowledgments. E.A.F. thanks the Zachry Mechanical Engineering Department Chair Endowment and the National Institutes of Health (Grant No. R01HL159300) for providing support for the execution of this work.

Disclosure of Interests. M.R., A.N.M.S., Z.Z., S.C.M., V.S.K., and E.A.F. declare that they have no competing interests. M.K.E. declares consulting relationships with W.L. Gore and Boston Scientific.

References

1. Chen, R.J., Ding, T., Lu, M.Y., et al.: Towards a general-purpose foundation model for computational pathology. Nature Medicine **30**, 850–862 (2024)
2. Fillinger, M.F., et al.: In vivo analysis of mechanical wall stress and abdominal aortic aneurysm rupture risk. Journal of Vascular Surgery **36**(3), 589–597 (2002)
3. Gu, Y., Saito, K., Ma, J.: Learning contrastive multimodal fusion with improved modality dropout for disease detection and prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 280–290. Springer (2025)
4. Han, X., Nguyen, H., Harris, C., Ho, N., Saria, S.: Fusemoe: Mixture-of-experts transformers for fleximodal fusion. Advances in Neural Information Processing Systems **37**, 67850–67900 (2024)
5. Hosny, A., Parmar, C., Coroller, T.P., et al.: Deep learning for lung cancer prognostication. PLOS Medicine **15**(11), e1002711 (2018)
6. Huang, S.C., Pareek, A., Seyyedi, S., Banerjee, I., Lungren, M.P.: Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. npj Digital Medicine **3**, 136 (2020)
7. Ilse, M., Tomczak, J.M., Welling, M.: Attention-based deep multiple instance learning. In: International Conference on Machine Learning (ICML). pp. 2127–2136 (2018)
8. Lambin, P., Leijenaar, R.T., Deist, T.M., et al.: Radiomics: The bridge between medical imaging and personalized medicine. Nature Reviews Clinical Oncology **14**(12), 749–762 (2017)
9. Liu, S., Zhang, B., Fang, R., Rueckert, D., Zimmer, V.A.: Dynamic graph neural representation based multi-modal fusion model for cognitive outcome prediction

in stroke cases. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 338–347. Springer (2023)
10. Liu, S., Zhang, B., Zimmer, V.A., Rueckert, D.: Multi-modal data fusion with missing data handling for mild cognitive impairment progression prediction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 293–302. Springer (2024)
11. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
12. Meng, Y., Huang, J., Wang, G., Zhang, C., Zhuang, H., Kaplan, L., Han, J.: Spherical text embedding. Advances in neural information processing systems **32** (2019)
13. Roby, M., Restrepo, J., Shan, D., Muluk, S., Eskandari, M., Kashyap, V., Finol, E.: An integrated framework for automated image segmentation and personalized wall stress estimation of abdominal aortic aneurysms. Bioengineering **13**(2), 191 (2026)
14. Roby, M., Sakib, A.N.M., Zhang, Z., Muluk, S.C., Eskandari, M.K., Finol, E.A.: Automatic explainable segmentation of abdominal aortic aneurysm from computed tomography angiography. IEEE Access **13** (2025)
15. Sakalihasan, N., Limet, R., Defawe, O.D.: Abdominal aortic aneurysm. The Lancet **365**(9470), 1577–1589 (2005)
16. Sakalihasan, N., Michel, J.B., Katsargyris, A., Kuivaniemi, H., Defraigne, J.O., Nchimi, A., Powell, J.T., Yoshimura, K., Hultgren, R.: Abdominal aortic aneurysms. Nature Reviews Disease Primers **4**, 34 (2018)
17. Shazeer, N., et al.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In: International Conference on Learning Representations (ICLR) (2017)
18. Susetzky, T., Qiu, H., Braren, R., Rueckert, D.: A holistic time-aware classification model for multimodal longitudinal patient data. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 24–33. Springer (2025)
19. Valanarasu, J.M.J., et al.: Medical transformer: Gated axial-attention for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 36–46. Springer (2021)
20. Wu, C., Shuai, Z., Tang, Z., Wang, L., Shen, L.: Dynamic modeling of patients, modalities and tasks via multi-modal multi-task mixture of experts. In: International Conference on Learning Representations (ICLR) (2025)
21. Yun, S., Choi, I., Peng, J., Wu, Y., Bao, J., Zhang, Q., Xin, J., Long, Q., Chen, T.: Flex-moe: Modeling arbitrary modality combination via the flexible mixture-of-experts. Advances in Neural Information Processing Systems **37**, 98782–98805 (2024)
22. Zhang, D., Li, Y., Zhang, Z.: Deep metric learning with spherical embedding. Advances in Neural Information Processing Systems **33**, 18772–18783 (2020)
23. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: Unet++: A nested u-net architecture for medical image segmentation. IEEE Transactions on Medical Imaging **39**(6), 1856–1867 (2020)