AS-TIME: Time-Conditioned Multimodal Modeling of Aortic Stenosis Progression

Diane Kim[1], Minh Nguyen Nhat To[2], Victoria Wu[1], Teresa S.M. Tsang[3], Christina Luong[3], and Purang Abolmaesumi[1]

[1] University of British Columbia, Vancouver, Canada
daeun96@student.ubc.ca, purang@ece.ubc.ca
[2] McGill University, Montreal, Canada
[3] Vancouver General Hospital, Vancouver, Canada

Abstract. Aortic stenosis (AS) is characterized by highly heterogeneous progression trajectories, making personalized surveillance and timely intervention challenging. We propose AS-TIME, a time-conditioned framework that predicts patient-specific AS progression from a single baseline echocardiogram and follow-up interval Δt. AS-TIME hierarchically integrates: **(i)** attention-based multiple instance learning to aggregate video and report representations, **(ii)** Δt-conditioned modality gating for adaptive multimodal fusion, and **(iii)** a Δt-guided mixture-of-experts head to model heterogeneous progression dynamics. AS-TIME achieves an AUC of 80.4% and balanced accuracy of 73.8%, outperforming all baselines. Temporal ablation demonstrate that explicit Δt conditioning improves both modality fusion and expert routing. The model provides multi-level explainability through modality fusion weights and expert utilization. Our code is available at: https://github.com/DeepRCL/AS-TIME.

Keywords: Temporal Modeling · Hierarchical Learning · Multimodal.

1 Introduction

Aortic stenosis (AS) is the most prevalent heart valve disease (HVD), associated with the narrowed opening of the aortic valve from calcification or fibrosis. AS is known for its life-threatening complications such as heart failure and stroke if left unmanaged [13]. With the global aging population and the rising burden of cardiovascular diseases, the prevalence of HVD continues to grow and is expected to increase substantially in the coming decades [1].

Despite the high prevalence, early detection and longitudinal disease monitoring remain suboptimal. No pharmacological therapy exists for AS, and surgical valve replacement is largely reserved for patients with advanced disease. Determining the optimal timing of surveillance and intervention — balancing the risks of premature replacement against the consequences of delayed treatment — remains a complex and nuanced clinical decision [6, 14].

As therapeutic control of disease progression is limited, it is often difficult to anticipate individual trajectories [12]. Some patients remain stable for years,

requiring only routine echocardiography (echo) surveillance, while others may deteriorate rapidly and necessitate earlier intervention [6]. This heterogeneity complicates the design of personalized follow-up schedules and monitoring intervals. Accurate prediction of progression is therefore critical for not only timely intervention, but also for optimizing healthcare resource allocation and reducing unnecessary testing or delayed referrals.

In the general medical domain, recent deep learning models have leveraged temporal encoding for longitudinal prediction tasks. Susetzky *et al.* (2025) [11] proposed a time-aware multimodal paradigm that encodes longitudinal electronic health record (EHR) data with temporal information to model entire patient trajectories for downstream prediction tasks. Here, temporal encoding is attached to each token at the feature space, providing rich metadata. Liu *et al.* (2025) [4] proposes a longitudinal multimodal framework where time embeddings are used as relative interval-aware alignment signals that guide cross-attention between the EHRs and chest X-ray images, enabling temporally-localized fusion for downstream disease prediction.

Despite the emergence of large-scale, time-aware prediction models, to the best of our knowledge, no prior work has explored these methods in longitudinal echo analysis, and in the context of this paper, specifically for predicting AS progression. Most clinical studies compare basic machine learning such as decision trees and logistic regression for predicting the occurrence of a clinical endpoint (e.g., valve replacement, mortality, or change in AS severity) using a comprehensive set of tabular data (e.g., clinical, laboratory, and echo measurements), some reaching near 200 individual features [3, 8]. Furthermore, these studies are trained to predict the occurrence of progression within a static time window (e.g., 5 years), lacking per-patient temporal modeling.

Contributions

1. In this paper, we propose **AS-TIME** (**T**emporal **I**nference **M**odel with **E**chocardiography), a time-conditioned multimodal framework that incorporates the inter-echo interval Δt to modulate multimodal fusion for patient-specific prediction of AS progression. Given a single baseline echo, AS-TIME estimates progression risk at a queried follow-up horizon. To the best of our knowledge, AS-TIME is the first to predict AS progression from a single baseline echo and elapsed time alone, without relying on additional clinical variables.
2. A hierarchical fusion architecture sequentially performs: **(i)** multiple instance learning (MIL)-based token aggregation, **(ii)** disagreement-aware modality gating, and **(iii)** Δt-conditioned mixture-of-experts (MoE) routing to adaptively model heterogeneous AS trajectories. In doing so, the model introduces Δt as a conditioning signal that adaptively modulates multimodal fusion and expert routing according to patient-specific prediction horizons.
3. By design, AS-TIME is inherently interpretable at multiple levels, exposing instance-level attention weights, modality contribution scores, and expert routing probabilities.

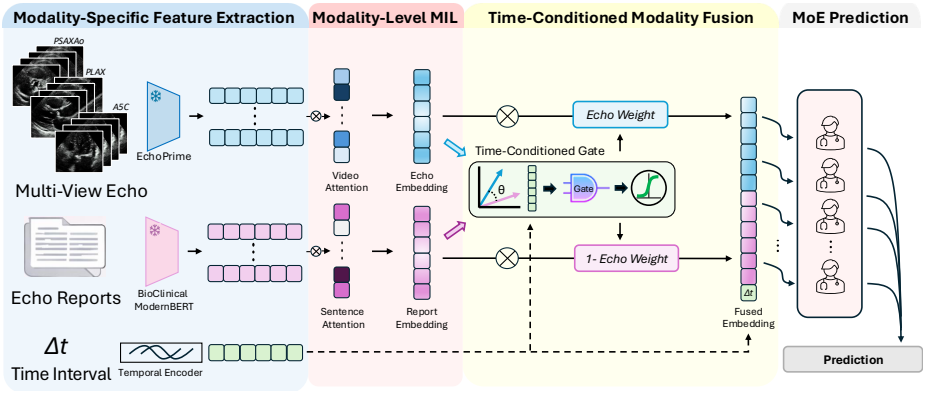

Fig. 1. Overview of the AS-TIME architecture. Echo videos and its report at t_0 are encoded with frozen EchoPrime [16] and BioClinical ModernBERT [10] respectively. Resulting embeddings are aggregated via token-level MIL, and a time-conditioned gating network fuses modalities using Δt. The final prediction y is derived from the weighted averaging of 12 expert predictions.

2 Methods

An overview of the AS-TIME architecture is outlined in Figure 1. The model learns a hierarchical mapping $f(x_{\text{vid}}, x_{\text{txt}}, \Delta t) \to y$, where x_{vid} denotes multi-view echo videos acquired at baseline t_0, x_{txt} is the associated clinical report, and Δt is the time interval to follow-up t_1. $y \in \{0, 1\}$ denotes the binary classification label, with $y_i = 0$ representing no AS progression and $y_i = 1$ representing AS progression at follow-up.

2.1 Modality-Specific Feature Extraction

For effective feature extraction, we used EchoPrime [16] and BioClinical ModernBERT [10] encoders for videos and reports, respectively. EchoPrime [16] is a state-of-the-art echo foundation model pretrained on over 12 million study-report pairs. BioClinical ModernBERT [10] is pretrained on 20 clinical datasets, leveraging its longer context window and increased vocabulary size. All input modalities were then projected onto a shared latent space for downstream fusion and prediction.

Echocardiogram Videos. Each study consists of videos $x_{\text{vid}} = \{x_1, \ldots, x_L\}$. A 512-dimensional feature is extracted for each video. The embeddings are then mapped through a learnable linear projection as $x_\ell \mapsto \mathbf{h}_\ell^{\text{vid}} \in \mathbb{R}^d$ with $d = 256$.

Echocardiogram Reports. Each clinical report, x_{txt}, is parsed into individual sentences where $x_{\text{txt}} = \{s_1, \ldots, s_S\}$. Each sentence is then encoded and projected as $s_s \mapsto \mathbf{h}_s^{\text{txt}} \in \mathbb{R}^d$.

Time Interval Encoding. The study interval Δt is log-transformed and standardized to obtain Δt_{std}. A sinusoidal embedding is then computed with frequencies $\{\omega_k\}_{k=1}^{K}$ as $\Delta t \mapsto \mathbf{h}^{\text{time}} \in \mathbb{R}^d$, where $2K = d$.

2.2 Time-Conditioned Modality Fusion

Token-Level Multiple Instance Learning. Within each modality, gated-attention MIL [2] is used to aggregate variable-length tokens into a study-level representation. For video tokens, importance weights are computed as:

$$\alpha_\ell^{\text{vid}} = \frac{\exp(g_{\text{vid}}(\mathbf{h}_\ell^{\text{vid}}))}{\sum_k \exp(g_{\text{vid}}(\mathbf{h}_k^{\text{vid}}))}, \tag{1}$$

to obtain the aggregated representation:

$$\mathbf{h}^{\text{vid}} = \sum_\ell \alpha_\ell^{\text{vid}} \mathbf{h}_\ell^{\text{vid}}. \tag{2}$$

The same attention-based aggregation is used for text tokens to compute $\mathbf{h}^{\text{txt}}$.

Modality-Level Gated Fusion. A cosine similarity is computed between $\mathbf{h}^{\text{vid}}$ and $\mathbf{h}^{\text{txt}}$, which then concatenates with $\mathbf{h}^{\text{time}}$ and passed through a gated fusion network g_{GF}. This produces the video weight $w_{\text{vid}} \in [0, 1]$. The report weight w_{txt} is defined as $1 - w_{\text{vid}}$. The final fused representation is derived by:

$$\mathbf{h}^{\text{fused}} = w_{\text{vid}} \mathbf{h}^{\text{vid}} + (w_{\text{txt}}) \mathbf{h}^{\text{txt}}. \tag{3}$$

2.3 Mixture-of-Experts Prediction

To model heterogeneous progression patterns, an MoE head with E experts is used. The gating network is a lightweight MLP that computes E weights as

$$\boldsymbol{\pi} = \text{softmax}\big(g_{\text{MoE}}([\mathbf{h}, \Delta t])\big). \tag{4}$$

Each expert produces logits $\mathbf{o}_e$, and the final prediction is a weighted averaging of all E experts: $\mathbf{o} = \sum_{e=1}^{E} \pi_e \mathbf{o}_e$.

3 Experiments

3.1 Datasets and Preprocessing

Our dataset included 2,605 AS patients who: **(i)** have undergone two or more full transthoracic echo studies at a tertiary medical centre between 2000 and 2020, **(ii)** study at t_0 showed early (i.e., mild) AS, and **(iii)** study at t_1 showed early (i.e., mild) or significant (i.e., moderate and severe) AS. All cases that underwent surgical intervention between t_0 and t_1 was removed. A total of 4,124

Table 1. Aortic stenosis progression prediction performance. '+' denotes embedding concatenation. Cross Attention used Δt and report as Q and video as KV. To demonstrate the performance difference between BERT and a medical LLM, 16-shot learning was performed with MedGemma-1.5-4B-IT [9]. Highest result is **bolded**, and the second highest result is underlined.

Method	Video	Report	AUC(%) ↑	BAcc(%) ↑	F1 ↑
Temporal Embedding			67.72 (0.01)	64.18 (0.65)	0.62 (0.00)
EchoPrime [16]+TempEmb	✓		77.17 (0.37)	70.46 (0.63)	0.70 (0.01)
ModernBERT [10]+TempEmb		✓	71.67 (0.18)	66.83 (1.37)	0.66 (0.01)
MedGemma-1.5-4B-IT [9]*		✓	-	56.34	0.52
TAMME [11]	✓	✓	65.11 (1.98)	61.52 (1.74)	0.60 (0.01)
Concatenation	✓	✓	78.09 (0.47)	71.20 (0.48)	0.70 (0.01)
Cross-Attention	✓	✓	79.06 (0.35)	71.96 (1.13)	0.71 (0.01)
AS-TIME (Ours)	✓	✓	**80.40 (0.23)**	**73.75 (0.48)**	**0.72 (0.01)**

*Raw scalar Δt used.

pairs were identified with a class distribution of 63.4% early AS ($y_i = 0$) and 36.6% significant AS ($y_i = 1$) at t_1.

A total of 3,760 apical-5-chamber (A5C), 15,598 parasternal-long axis (PLAX) and 16,260 parasternal-short axis aortic valve (PSAXAo) videos were extracted. The views were specifically chosen for containing AS diagnostic information [5, 7]. 32 frames were sampled with a stride of 2, normalized and resized to 224 × 224 pixels. Dataset was split at the patient level into training, validation, and test sets containing 2,933, 594, and 597 pairs respectively, ensuring that all pairs from a given patient were assigned to a single split. Class distributions (approximately 64% $y_i = 0$ and 36% $y_i = 1$) and mean follow-up interval Δt (2.35, 2.46, and 2.46 years) were preserved across splits. All results are reported on the held-out test set as the mean of five random seeds, with hyperparameters selected using the validation set.

3.2 Implementation Details

The model was trained using the AdamW optimizer (lr 3.8×10^{-6}, weight decay 10^{-3}) for 50 epochs with batch size 16. To address class imbalance, we sample the training set with probability inversely proportional to class frequency. The MoE module uses 12 experts. All baselines were optimized to the best of our abilities.

4 Results and Discussion

4.1 Baselines Comparison

Table 1 summarizes the comparative performance of AS-TIME against the baselines. Δt (TempEmb) alone was first evaluated to quantify its standalone contri-

Table 2. Temporal modeling ablation. **GF** denotes Gated Fusion (modality weighting), and **MoE** denotes Mixture-of-Experts (final classification). Highest result is **bolded**, and the second highest result is underlined.

Δt				
GF	**MoE**	**AUROC(%) ↑**	**BAcc(%) ↑**	**F1 ↑**
		73.91 (0.43)	67.98 (0.93)	0.67 (0.00)
✓		79.42 (1.26)	72.49 (2.00)	0.71 (0.02)
✓	✓	**80.40 (0.23)**	**73.75 (0.48)**	**0.72 (0.01)**

bution. TempEmb achieved an AUROC of 67.72% and BAcc of 64.18%, indicating that time horizon partly carries predictive signal. This is clinically aligned, as a longer time horizon increases the chance of AS progression. However, it also demonstrates that time alone cannot explain disease progression. The predictive performances of video and reports with Δt were then evaluated independently. EchoPrime [16]+TempEmb showed the highest performance of AUC 77.17% and BAcc 70.46%. MedGemma-1.5-4B-IT [9] with raw scalar Δt showed poorer prediction accuracy (BAcc 56.34%) at 16-shot learning than BioClinical ModernBERT[10]+TempEmb.

Incorporation of both modalities demonstrated prediction performance improvement. Cross-Attention showed the highest performance gain (AUC 79.06%, BAcc 71.96%), followed by Concatenation (AUC 78.09%, BAcc 71.20%). TAMME [11] underperformed relative to other multimodal approaches, likely due to the model requiring multiple input data across multiple time points. This further enhances our model's strength and data efficiency in using only a single time point echo study to provide a personalized, time progression-aware AS trajectory.

Finally, we observe that AS-TIME achieves the highest performance across all metrics. The model displays up to 15.29% increase in AUC, 17.34% increase in BAcc, and 0.2 increase in F1. This indicates that AS-TIME's multimodal and hierarchical information handling provides measurable and consistent improvements. Relative to the strongest baseline (Cross-Attention), these gains were consistent across all five seeds, with paired t-tests confirming statistically significant improvements ($p < 0.05$).

4.2 Ablation Study

To evaluate the effect of Δt signaling in multimodal fusion, an ablation study was conducted where Δt was marked as zero (Table 2). Specifically, two variants from AS-TIME was explored: **(i)** no temporal signaling in both gated fusion (GF) and MoE, and **(ii)** temporal signaling in GF only.

The results show AS-TIME's superior performance against TAMME [11] at full Δt ablation, suggesting that the hierarchical information fusion of the model itself provides representational advantages even without explicit temporal information. With the introduction of Δt in gated fusion, however, the performance

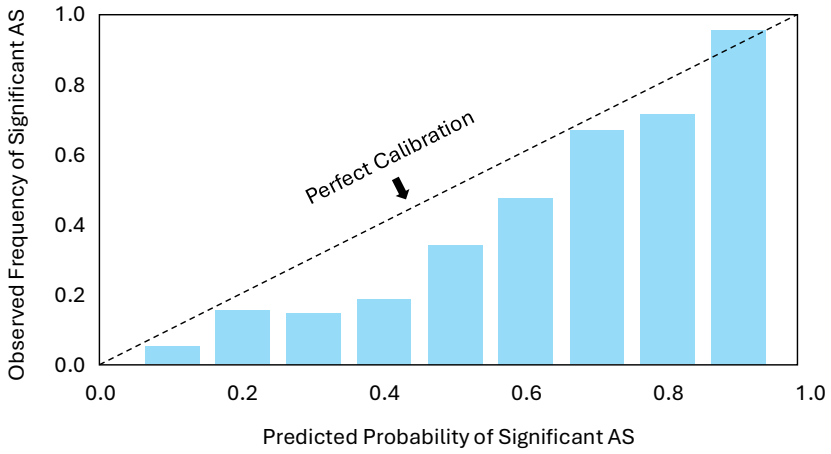

Fig. 2. AS-TIME Prediction Reliability diagram. The x-axis is the model's predicted probability of significant AS at t_1 and the y-axis is the ground truth observed frequency of significant AS at each predicted probability. Dashed line denotes perfect calibration.

improves substantially - 5.51% in AUC, 4.42% in BAcc, and 0.04 in F1. This performance gain suggests that Δt serves as an effective conditioning signal for adaptive video-report fusion. Conditioning expert routing on Δt yields additional performance gains, indicating that Δt provides complementary information at both the representation and decision levels. The reduced variance across seeds imply that temporal conditioning likely stabilizes model performance and reduces sensitivity to initialization.

4.3 Explainability

Video-Report Weighting. At the gated fusion step, predominantly negative cosine similarities were observed between the aggregated video and report embeddings, with an average of -0.33. This disagreement suggests that: **(i)** MIL attends to different information contained in videos and reports, and **(ii)** the MIL representations offer complementary information that result in improved performance against either modality alone (Table 1).

There is a strong reliance on video where the average video weight is 0.96 - this confirms that the videos serve as the primary predictive signal and report provides secondary contextual cues where appropriate, in line with the stronger performance of EchoPrime [16]+TempEmb over BioClinical ModernBERT [10]+TempEmb. This does not suggest that report features are uninformative, as multimodal baselines with videos and text yield higher results than EchoPrime [16]+TempEmb alone (Table 1). Further analysis is needed to fully characterize each modality's contributions.

Expert Weighting. The model design informs the routing trends of the MoE head, offering transparency into how the model comes to its final predictions. Different patterns were observed with experts specializing in specific Δt intervals. For example, the MoE head assigns Expert 8 a weight of 0.20 for Δt up to 8 years, decreasing to 0.10 thereafter. Expert 2's weight increases with Δt, peaking at 0.25 at 8-10 years before gradually declining. Expert 3 shows a steady increase with greater Δt, rising from 0.05 at 0-2 years to 0.17 at 14-16 years. Conversely, Expert 12 is most active in early intervals and progressively diminishes toward zero at longer follow-up periods. Not only does this demonstrate the model's adaptive prediction while preserving diversity among experts, it also highlights its interpretability and transparent progression dynamics. Across Δt bins, AUROC remained stable while BAcc showed a modest decrease at longer horizons, likely reflecting smaller subgroup sizes and the greater difficulty of longer-range prediction.

Model Uncertainty. AS-TIME is well-calibrated, with predicted probabilities closely matching ground-truth frequencies (Figure 2). Although MoE architectures do not inherently guarantee improved calibration [15], conditional routing on Δt may promote expert diversity that stabilizes probability estimates through ensemble-like averaging.

4.4 Clinical Relevance and Limitations

AS-TIME demonstrates that explicit conditioning on elapsed time significantly improves multimodal fusion and prediction of AS progression. By treating Δt as a structural signal, the model reshapes how echo videos and reports are interpreted, moving beyond static classification toward trajectory-aware inference. A set of echo findings may imply different risks depending on progression horizon, and integrating Δt enables patient-specific risk estimation. This supports individualized follow-up decisions, helping clinicians prioritize patients who require earlier reassessment while reducing unnecessary surveillance for stable cases. More importantly, AS-TIME operates using only a single baseline echo study, without additional demographic or clinical variables. In resource-limited settings, including rural and remote communities with constrained imaging and referral capacity, such adaptive risk stratification may help allocate care more efficiently while maintaining patient safety.

A key limitation is the lack of external validation. The study is based on data from a single institution, and no publicly available cohort currently provides longitudinal echo studies, clinical reports, and standardized AS severity labels required for independent evaluation. Validation across multiple institutions is therefore necessary to assess the model's generalizability. Future work should also extend AS-TIME through rigorous subgroup analyses across clinically relevant factors such as age, smoking status, and family history, examining both predictive performance and decision patterns across distinct patient phenotypes. Expanding the dataset to include patients with three or more longitudinal

echo studies will enable richer modeling of progression trajectories. An additional downstream direction is to invert the framework to estimate individualized time-to-progression (Δt), further enhancing its clinical utility.

5 Conclusion

In this paper we introduced AS-TIME, a time-conditioned multimodal framework for patient-specific prediction of AS progression. By structurally integrating elapsed time into hierarchical fusion, the model achieved an AUC of 80.40% and balanced accuracy of 73.75%, outperforming all baselines while explicitly decomposing prediction into instance selection, modality weighting, and trajectory-conditioned expert routing. Beyond performance gains, AS-TIME reframes progression modeling as time-aware representation learning rather than static feature extraction and classification. Time-conditioned architectures such as AS-TIME may act as a foundation for precision surveillance in HVD, enabling adaptive follow-up strategies and advancing trajectory-aware clinical modeling.

Acknowledgments. This research was supported in part by the Canadian Institutes of Health Research (CIHR) and the Natural Sciences and Engineering Research Council of Canada (NSERC), and through computational resources and services provided by Advanced Research Computing at the University of British Columbia.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chen, Q., Shi, S., Wang, Y., Shi, J., Liu, C., Xu, T., Ni, C., Zhou, X., Lin, W., Peng, Y., Zhou, X.: Global, regional, and national burden of valvular heart disease, 1990 to 2021. Journal of the American Heart Association **13**(24), e037991 (2024)
2. Ilse, M., Tomczak, J., Welling, M.: Attention-based deep multiple instance learning. In: Proceedings of the 35th International Conference on Machine Learning. vol. 80, pp. 2127–2136. PMLR (2018)
3. Itelman, E., Shapira, Y., Shechter, A., Loebl, N., Altman, Y., Perl, L., Kornowski, R.: Prediction of aortic stenosis progression using artificial intelligence: A machine learning model. JACC: Advances **4**(10_Part_2), 102121 (2025)
4. Liu, C., Yao, W., Yin, K., Cheung, W.K., Qin, J.: Multimodal disease progression modeling via spatiotemporal disentanglement and multiscale alignment. In: Advances in Neural Information Processing Systems. vol. 38, pp. 132539–132566. Curran Associates, Inc. (2025)
5. Manzo, R., Ilardi, F., Nappa, D., Mariani, A., Angellotti, D., Molaro, M.I., Sgherzi, G., Castiello, D.S., Simonetti, F., Santoro, C., Canonico, M.E., Avvedimento, M., Piccolo, R., Franzone, A., Esposito, G.: Echocardiographic evaluation of aortic stenosis: A comprehensive review. Diagnostics **13**(15), 2527 (2023)
6. Otto, C.M., Nishimura, R.A., Bonow, R.O., Carabello, B.A., Erwin III, J.P., Gentile, F., Jneid, H., Krieger, E.V., Mack, M., McLeod, C., O'Gara, P.T., Rigolin,

V.H., Sundt, T.M., Thompson, A., Toly, C.: 2020 acc/aha guideline for the management of patients with valvular heart disease: A report of the american college of cardiology/american heart association joint committee on clinical practice guidelines. Circulation **143**(5), e72–e227 (2021)
7. Pazos-López, P., Paredes-Galán, E., Peteiro-Vázquez, J., López-Rodríguez, E., García-Rodríguez, C., Bilbao-Quesada, R., Blanco-González, E., González-Ríos, C., Calvo-Iglesias, F., Íñiguez-Romo, A.: Value of non-apical echocardiographic views in the up-grading of patients with aortic stenosis. Scandinavian Cardiovascular Journal **55**(5), 279–286 (2021)
8. Sanabria, M., Tastet, L., Pelletier, S., Leclercq, M., Ohl, L., Hermann, L., Mattei, P.A., Precioso, F., Coté, N., Pibarot, P., Droit, A.: Ai-enhanced prediction of aortic stenosis progression: Insights from the progressa study. JACC: Advances **3**(10), 101234 (2024)
9. Sellergren, A., Kazemzadeh, S., Jaroensri, T., Kiraly, A., Traverse, M., Kohlberger, T., Xu, S., Jamil, F., Hughes, C., Lau, C., et al.: Medgemma technical report (2025)
10. Sounack, T., Davis, J., Durieux, B., Chaffin, A., Pollard, T.J., Lehman, E., Johnson, A.E., McDermott, M., Naumann, T., Lindvall, C.: Bioclinical modernbert: A state-of-the-art long-context encoder for biomedical and clinical nlp (2025)
11. Susetzky, T., Qiu, H., Braren, R., Rueckert, D.: A holistic time-aware classification model for multimodal longitudinal patient data. In: Proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. vol. LNCS 15960, pp. 24–33. Springer Nature Switzerland (2025)
12. Takeji, Y., Tada, H., Taniguchi, T., Sakata, K., Kitai, T., Shirai, S., Takamura, M.: Current management and therapy of severe aortic stenosis and future perspective. Journal of Atherosclerosis and Thrombosis **31**(10), 1353–1364 (2024)
13. Thoenes, M., Bramlage, P., Zamorano, P., Messika-Zeitoun, D., Wendt, D., Kasel, M., Kurucova, J., Steeds, R.P.: Patient screening for early detection of aortic stenosis (as)—review of current practice and future perspectives. Journal of Thoracic Disease **10**(9), 5584–5594 (2018)
14. Vahanian, A., Beyersdorf, F., Praz, F., Milojevic, M., Baldus, S., Bauersachs, J., Capodanno, D., Conradi, L., De Bonis, M., De Paulis, R., Delgado, V., Freemantle, N., Gilard, M., Haugaa, K.H., Jeppsson, A., Jüni, P., Pierard, L., Prendergast, B.D., Sádaba, J.R., Tribouilloy, C., Wojakowski, W., ESC/EACTS Scientific Document Group, ESC National Cardiac Societies: 2021 esc/eacts guidelines for the management of valvular heart disease: Developed by the task force for the management of valvular heart disease of the european society of cardiology (esc) and the european association for cardio-thoracic surgery (eacts). European Heart Journal **43**(7), 561–632 (2022)
15. Verma, R., Barrejón, D., Nalisnick, E.: Learning to defer to multiple experts: Consistent surrogate losses, confidence calibration, and conformal ensembles. In: International Conference on Artificial Intelligence and Statistics. vol. 206, pp. 11415–11434. PMLR (2023)
16. Vukadinovic, M., Chiu, I.M., Tang, X., Yuan, N., Chen, T.Y., Cheng, P., Li, D., Cheng, S., He, B., Ouyang, D.: Comprehensive echocardiogram evaluation with view primed vision language ai. Nature **650**(8103), 970–977 (2026)