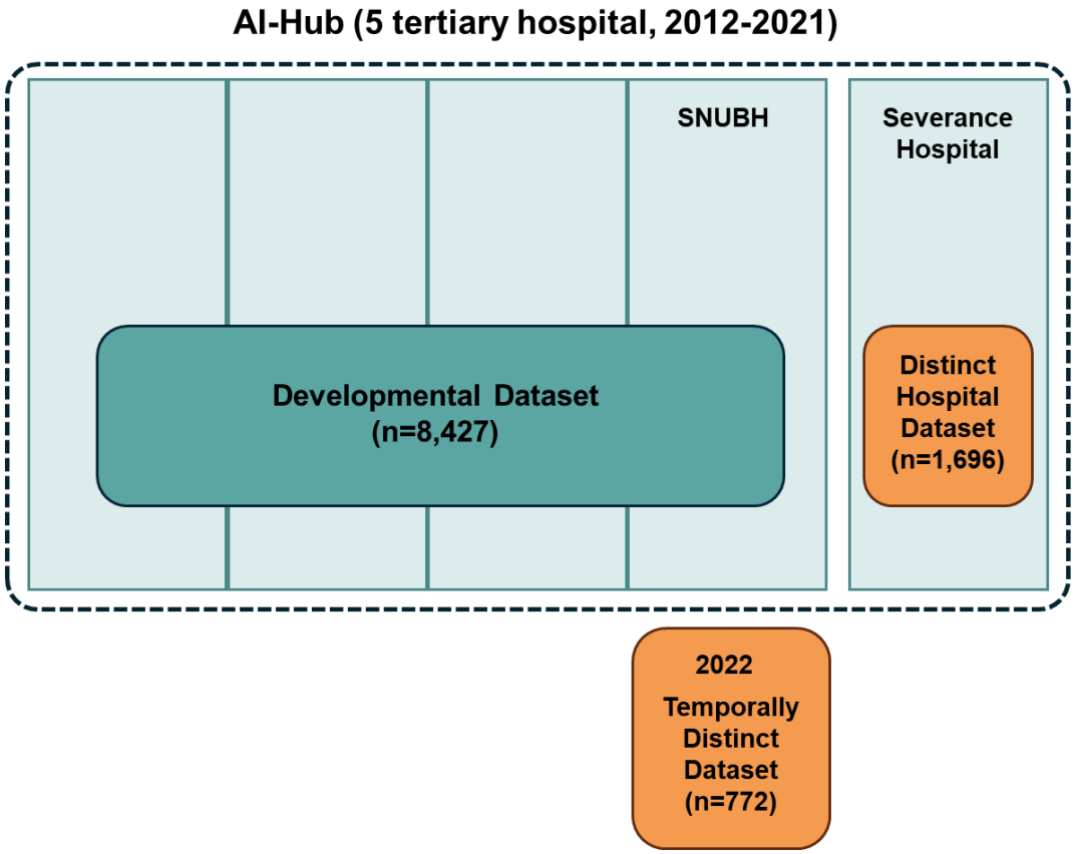
Supplemental Methods 1. Study Population

AI-Hub dataset consisted of 30,000 echocardiographic examinations retrospectively collected from five tertiary hospitals, including Chungnam National University Hospital, Hanyang University Hospital, Seoul National University Bundang Hospital, Severance Hospital, and Soonchunhyang University Seoul Hospital, over the period from 2012 to 2021. It encompasses a wide range of cardiovascular disease categories, from normal cases to ischemic heart disease, cardiomyopathy, pulmonary hypertension and embolism, pericardial disease, valvular heart disease, cardiac mass, and congenital heart disease.



The AI-based frameworks introduced here were all developed using data extracted from the AI-Hub dataset. Specifically, the DL-based AVS continuum assessment algorithm was developed using the **Developmental Dataset (DDS)** sourced from the AI-Hub. During the

assembly of DDS, data from Severance Hospital were deliberately excluded and used for external validation. We initially screened transthoracic echocardiography (TTE) data from 4,563 patients diagnosed with aortic valve stenosis (AVS). After excluding those who had undergone aortic valve (AV) replacement or open-heart surgery, those with moderate or greater AV regurgitation, or cases where the severity of AVS could not be determined, 4,018 AVS patients have remained. To ensure the model's accuracy, 628 cases exhibiting discordant findings among aortic valve (AV) peak velocity (V_{max}), mean pressure gradient (mPG), and aortic valve area (AVA) regarding the severity of AVS were excluded. These cases were later used separately for validating the model. Among the 3,390 AVS-diagnosed patients included in the model development, 2,500, 516, and 374 were categorized into mild, moderate, and severe AVS, respectively. Additionally, for the purpose of model training, TTE examinations were extracted for 3,290 individuals demonstrating normal AV morphology and function and 1,747 individuals exhibiting signs of AV sclerosis—characterized by degenerative changes in the AV but not meeting the diagnostic criteria for AVS. Consequently, TTE data from a total of 8,427 individuals were compiled into the DDS. These data were split in an 8:1:1 ratio for training, validation, and internal testing purposes.

The **Distinct Hospital Dataset (DHDS)** was compiled by reviewing data from Severance Hospital that were not included in the DDS sourced from the AI-Hub dataset. A total of 719 AVS patients were reviewed, none of whom had undergone AV replacement or open-heart surgery. After excluding 60 patients with moderate or greater AV regurgitation, the dataset included 659 AVS patients (209 mild, 251 moderate, and 199 severe). Adding 1,037 normal patients, the DHDS totaled 1,696 patients. Since Severance Hospital does not commonly use the diagnosis of AV sclerosis, a separate AV sclerosis category was not included.

For the **Temporally Distinct Dataset (TDDS)**, we screened TTE data conducted in

2022 at Seoul National University Bundang Hospital, identifying 520 eligible patients with AVS. After excluding cases with a documented history of AV replacement or open-heart surgery, those with moderate or greater AV regurgitation, or cases where the severity of AVS could not be determined, a total of 443 patients with AVS remained (313 mild, 75 moderate, and 55 severe). Additionally, 55 individuals with normal AV and 274 with AV sclerosis, identified during the same period, were included, resulting in a total of 772 cases included for this dataset.

Supplemental Methods 2. View Classification Networks Update

To enhance our echocardiographic view classification algorithm, we expanded our datasets and refined the classification algorithm to include new views. Building upon our previous work, which utilized 67,553 data points, the dataset was broadened to encompass more granular classifications and additional views.[1] Specifically, we differentiated the parasternal long-axis (PLAX) zoomed view into four distinct categories: PLAX zoomed aortic valve (AV) (779 videos), PLAX zoomed mitral valve (MV) (279 videos), PLAX zoomed both AV and MV (1,357 videos), and PLAX zoomed aorta (502 videos). Additionally, we incorporated 663 CW Doppler AV images obtained from the right parasternal view. These enhancements are crucial for accurately measuring AV stenosis (AVS) parameters, such as left ventricular outflow tract (LVOT) diameter, AV maximum velocity (V_{max}), mean pressure gradient (mPG), and AV area (AVA), which are critical for evaluating the severity of AVS. The new data points were annotated using a Developmental Dataset (DDS), and the view classification network was subsequently retrained with this enriched dataset, employing the methodology previously detailed in our research.[1]

Target Views for Current Version of View Classification Algorithm.

B-mode	Parasternal long-axis left ventricle
	Parasternal long-axis zoomed AV
	Parasternal long-axis zoomed MV
	Parasternal long-axis zoomed AV & MV
	Parasternal long-axis zoomed aorta
	Parasternal short-axis, level of great vessels
	Parasternal short-axis, level of mitral valve
	Parasternal short-axis, level of papillary muscle
	Parasternal short axis, level of apex
	Apical four-chamber
	Apical four-chamber zoomed left ventricle
	Apical four-chamber right ventricular-focused
	Apical five-chamber
	Apical two-chamber
	Apical two-chamber zoomed left ventricle
	Apical three-chamber
	Apical three-chamber zoomed left ventricle
	Subcostal four-chamber
	Subcostal long axis IVC
M-mode	M-mode through left ventricle
	M-mode through aorta and left atrium
	M-mode tricuspid annular plane systolic excursion
Spectral and tissue Doppler	PW Doppler mitral valve
	TDI mitral valve lateral annulus
	TDI mitral valve septal annulus
	CW Doppler mitral stenosis
	CW Doppler mitral regurgitation
	PW Doppler left ventricular outflow tract
	CW Doppler aortic valve
	CW Doppler aortic valve in parasternal
	CW Doppler aortic regurgitation
	CW Doppler tricuspid regurgitation
	PW Doppler right ventricular outflow tract
	CW Doppler pulmonic valve
	CW Doppler pulmonic regurgitation
	Pulmonary vein flow

AV, aortic valve; CW, continuous wave; IVC, inferior vena cava; MV, mitral valve; PW, pulse wave; TDI, tissue Doppler imaging

Supplemental Methods 3. DL-based AVS Continuum Assessment Algorithm

Given an input video $\boldsymbol{x}$, we extract a feature $\boldsymbol{z}$ using the r2plus1d architecture.[2] Importantly, we modified the backbone network to avoid temporal down-sampling by maintaining a stride of 1 along the temporal axis. From this feature $\boldsymbol{z}$, we employ four decoders, three of which are designed to predict continuous variables such as AV V_{max}, mPG, and AVA. Each auxiliary decoder is constructed to predict these continuous variables through regression. Each auxiliary decoder consists of two hidden layers with 512 units each, batch normalization and ReLU activation, followed by an output layer with a single unit and a Softplus activation function to ensure positive output values.

We then implemented a fusion module, termed the Adaptive Feature Fusion Module (AFFM), that generates a score for each continuous feature and then fuses them into one feature by a weighted sum of all features according to their scores. For features $z_{v_{max}}, z_{mPG}, z_{AVA}$ from the three auxiliary decoders before the output layer with a single unit, the fusion module computes a weight $w_{v_{max}}, w_{mPG}, w_{AVA}$ for each feature $z_{v_{max}}, z_{mPG}, z_{AVA}$ using a fully connected layer followed by batch normalization, ReLU activation, and a final fully connected layer with sigmoid activation. The fused feature z^{fused} is then computed as a weighted sum of the features: $z_{fused} = w_{v_{max}} z_{v_{max}} + w_{mPG} z_{mPG} + w_{AVA} z_{AVA}$

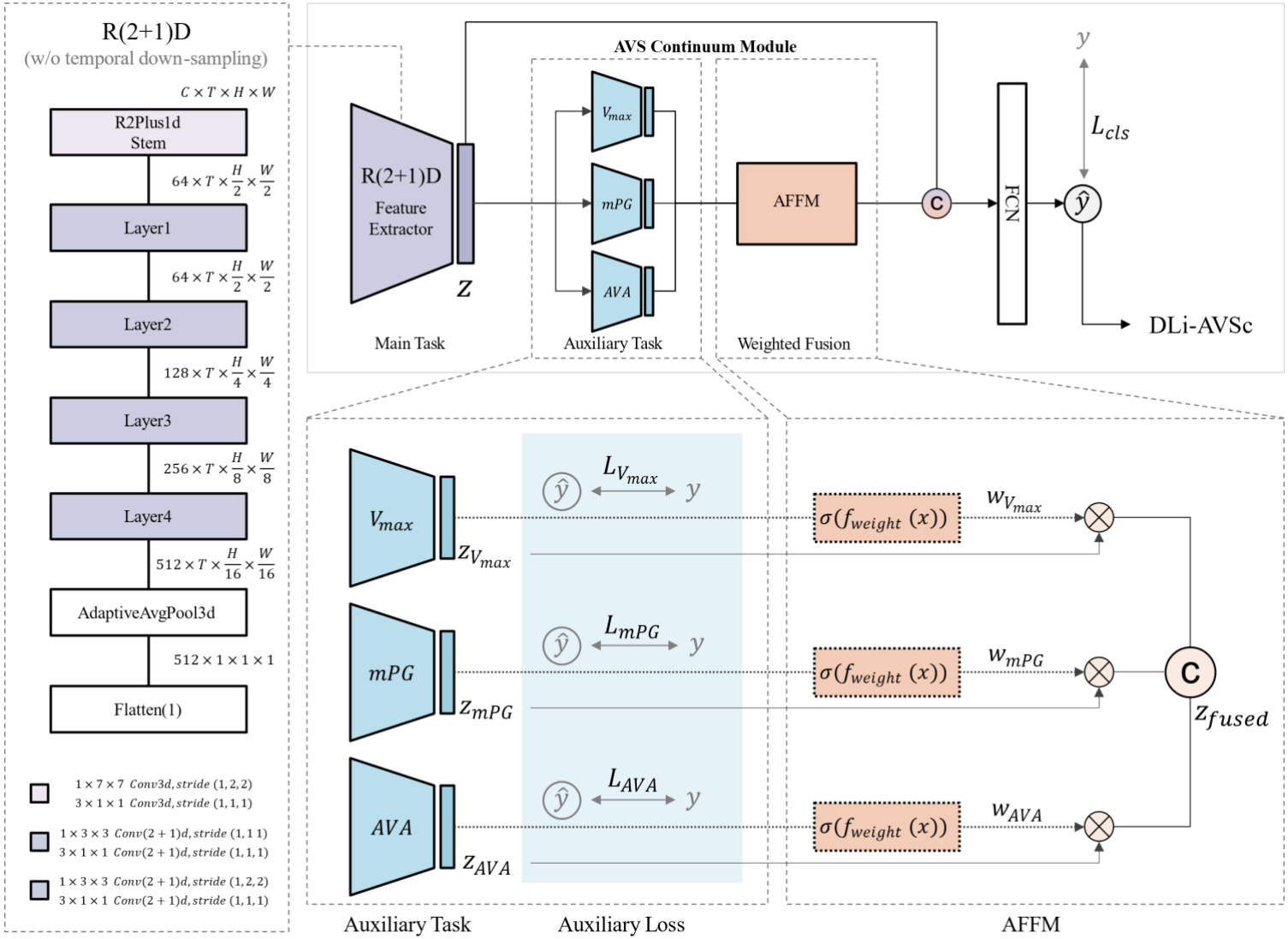
The fused feature z^{fused} is then concatenated with the original feature z from the r2plus1d backbone, resulting in the final feature z_{final}: $[z; z_{fused}]$. The final classifier classifier processes this concatenated feature z^{final} by linear function followed by a sigmoid function: $\widehat{y_{final}} = sigmoid(W z_{final} + b)$.

Input videos were resized to 224 x 224, and normalization was applied to a [0, 1] range for both training and inference. During training, if the number of frames exceeded the specified

clip length (16 frames), a random start index was selected. If the number of frames was fewer, indices were evenly spaced to fit the clip length. For inference, the videos were divided into four intervals and stacked. If the frames were insufficient for multiple intervals but exceeded the clip length, the video was sampled, and the clip was replicated to match the intervals. If the frames were fewer than the clip length, indices were evenly spaced, and the clip was duplicated to match the intervals. The stacked inputs from the inference phase are averaged to enhance the model's generalization capabilities. The final DL index for AVS continuum (DLi-AVSc) was rescaled to 0-100 by multiplying the model output after the sigmoid function by 100.

During training, a class sampler was used to balance the sampling of normal and AVS classes by assigning higher weights to the less frequent class. The Adam optimizer was used with a learning rate of 0.0001 and a batch size of 28. No learning rate scheduler was utilized. Early stopping was implemented with a tolerance of 300 epochs, monitoring the validation loss as the metric.

Deep Learning Architecture for DLi-AVSc

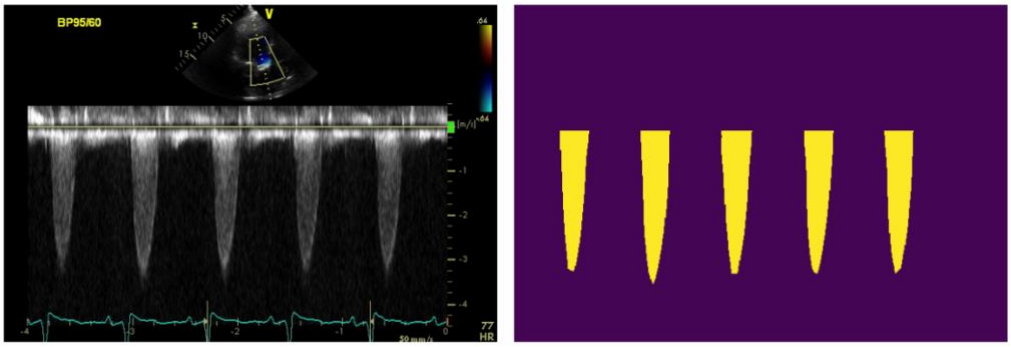


Supplemental Methods 4. Automated Conventional AVA Assessment Algorithm

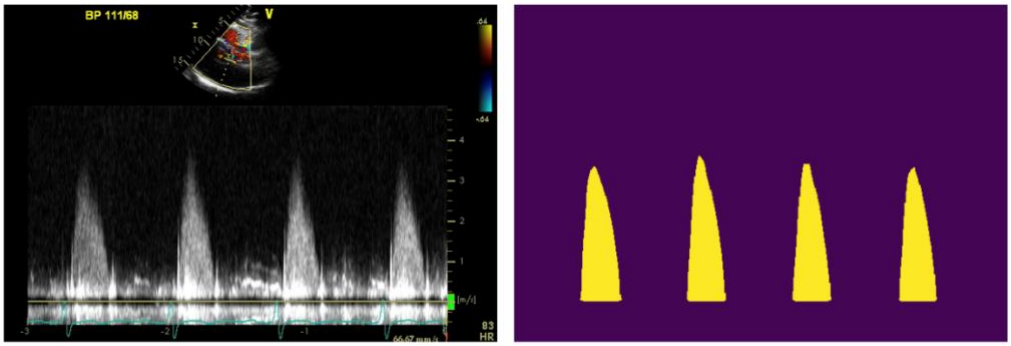
4.1 Automatic Measurement of Spectral Doppler Echocardiography

The Doppler segmentation network employs the BiSeNetV2 architecture, which has been thoroughly described in our previous publication.[3-5] The BiSeNetV2 is specifically designed to balance accuracy and computational efficiency, making it suitable for real-time applications. During the training process, the network was optimized using standard cross-entropy loss. Notably, our Doppler segmentation network did not include training data for the AV continuous wave (CW) Doppler from the right parasternal view. However, this view can be inferred as a vertically flipped version of the CW Doppler AV in the apical view. During inference, we preprocess the input by flipping the image vertically before running it through the network, allowing us to utilize the same trained model for both views without additional training data.[5]

AV CW Doppler from apical and right parasternal view



AV CW Doppler from Apical View

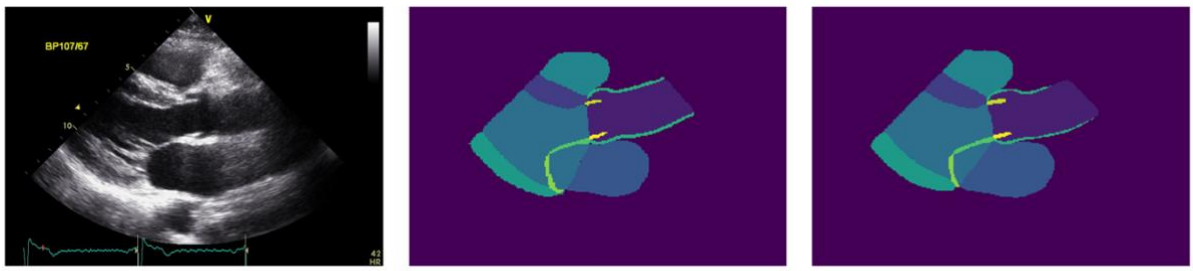


AV CW Doppler from Right Parasternal View

4.2 Segmentation Network for Parasternal Long-Axis (PLAX) View

For the PLAX segmentation network, we utilized the SegFormer architecture, which includes a transformer encoder that provides multiscale features without needing positional encoding and a lightweight multi-layer perceptron (MLP) decoder integrating local and global attention for efficient segmentation.[6] A weighted cross-entropy loss was applied during training to account for the relatively small size of the mitral valve (MV) and AV in the PLAX view. A total of 2,369 PLAX videos were annotated by experienced sonographers. The images were resized into 512 x 512 and normalized to [-1,1]. We used the Adam optimizer with a learning rate of 0.001 and incorporated RandAug, enhanced with echocardiography-specific augmentations such as shadow, depth attenuation, and haze, to improve model robustness.[7,8] Additionally, a cosine annealing learning rate schedule was employed to optimize the training process.[9] Complete videos are presented in **Video S1**.

Human Expert Annotation and AI Predicted Mask in PLAX View



Supplemental Methods 5. Quantifying Uncertainty in Predicted Segmentation

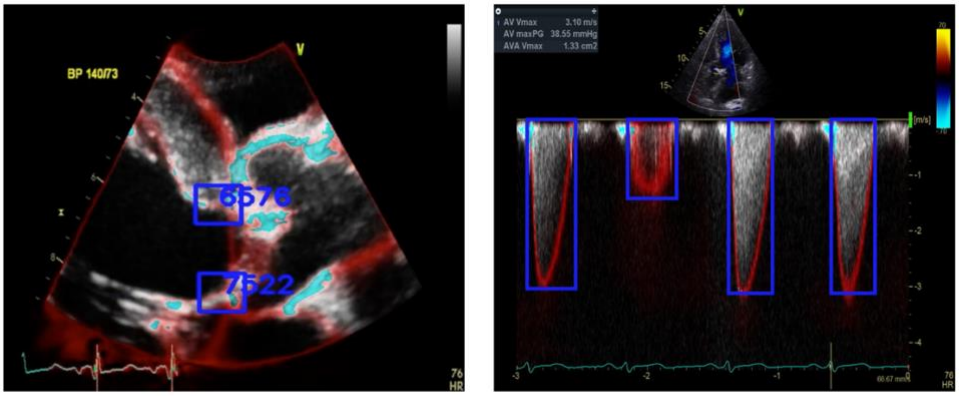
Quantifying uncertainty in segmentation predictions is a meticulous process, crucial because segmentation errors can significantly impact the accuracy of subsequent automatic measurements. This uncertainty arises from two primary sources: epistemic uncertainty, which arises from a lack of knowledge of the DL model, and aleatoric uncertainty, which results from inherent noise in the data. To quantify these uncertainties, we calculate predictive entropy from the segmentation network's probability map, which provides a measure of the total uncertainty by combining both aleatoric and epistemic uncertainties. The entropy is computed for each pixel in the segmentation map, allowing us to identify regions with high uncertainty. The entropy is calculated using the equation:

$$H[\boldsymbol{p_{ij}}] = \sum_c p_{ij}^c \log p_{ij}^c,$$

where i, j represent pixel coordinates and c represents the class.

For quantifying uncertainty in the LVOT measurements in the PLAX view, we focus on regions of interests (ROIs) that directly affect the performance of LVOT measurement. Using the detected two points marking the annulus, we set a 50 × 50 ROI (10% of the resized image) centered on these points, as shown in the figure below**.** We then summed the entropy of each pixel within this ROI to assess uncertainty. For Doppler measurements, we evaluated uncertainty for the Doppler signal in each single beat. With the detected significant Doppler flow, we create an ROI and crop the entropy map to the corresponding ROI, normalizing it to 64 x 64. By summing the entropy values within the normalized ROI, we obtain the quantified uncertainty for each Doppler signal by beat. Complete videos are presented in **Video S2**.

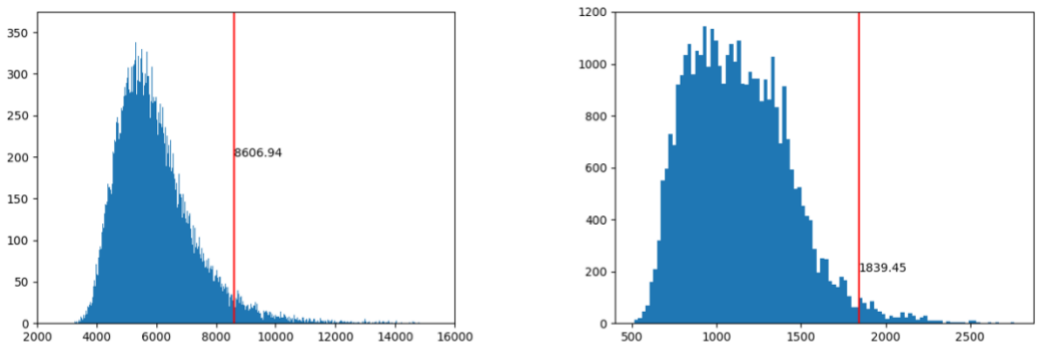
Regions of Interest marked for assessing uncertainty



PLAX zoomed AV CW Doppler AV

From the validation set used for training, we find thresholds to reject the frames or beats by identifying the top percentage of frames or beats with the highest uncertainty. Specifically, we reject the top 5% of frames with the highest uncertainty from PLAX AV zoomed videos and the top 1% of beats with the highest uncertainty from Doppler images. We calculate the mean of entropy scores and add 1.96 × s.d. for PLAX frames and 2.33 × s.d. for Doppler images. The distributions of the entropy scores for both PLAX and Doppler images are shown in the figure below, illustrating how the thresholds are set.

Distribution of Entropy Scores

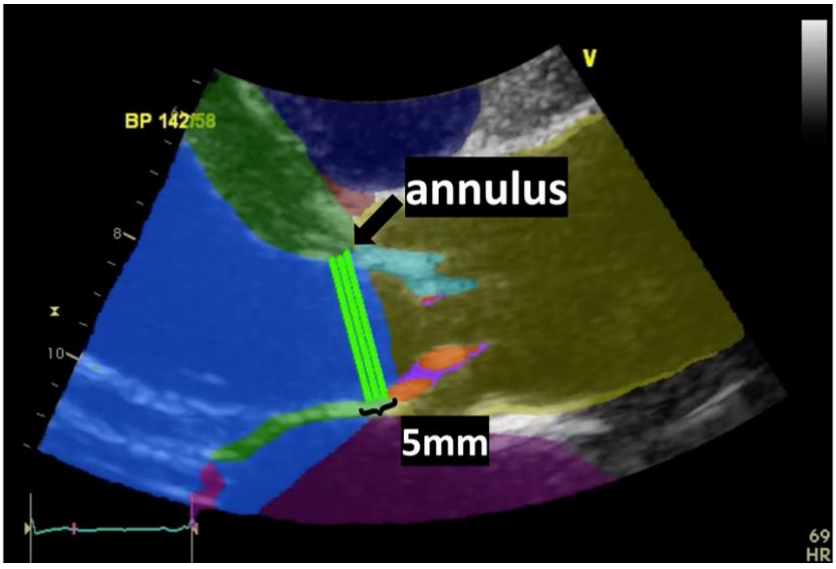


PLAX zoomed AV Doppler

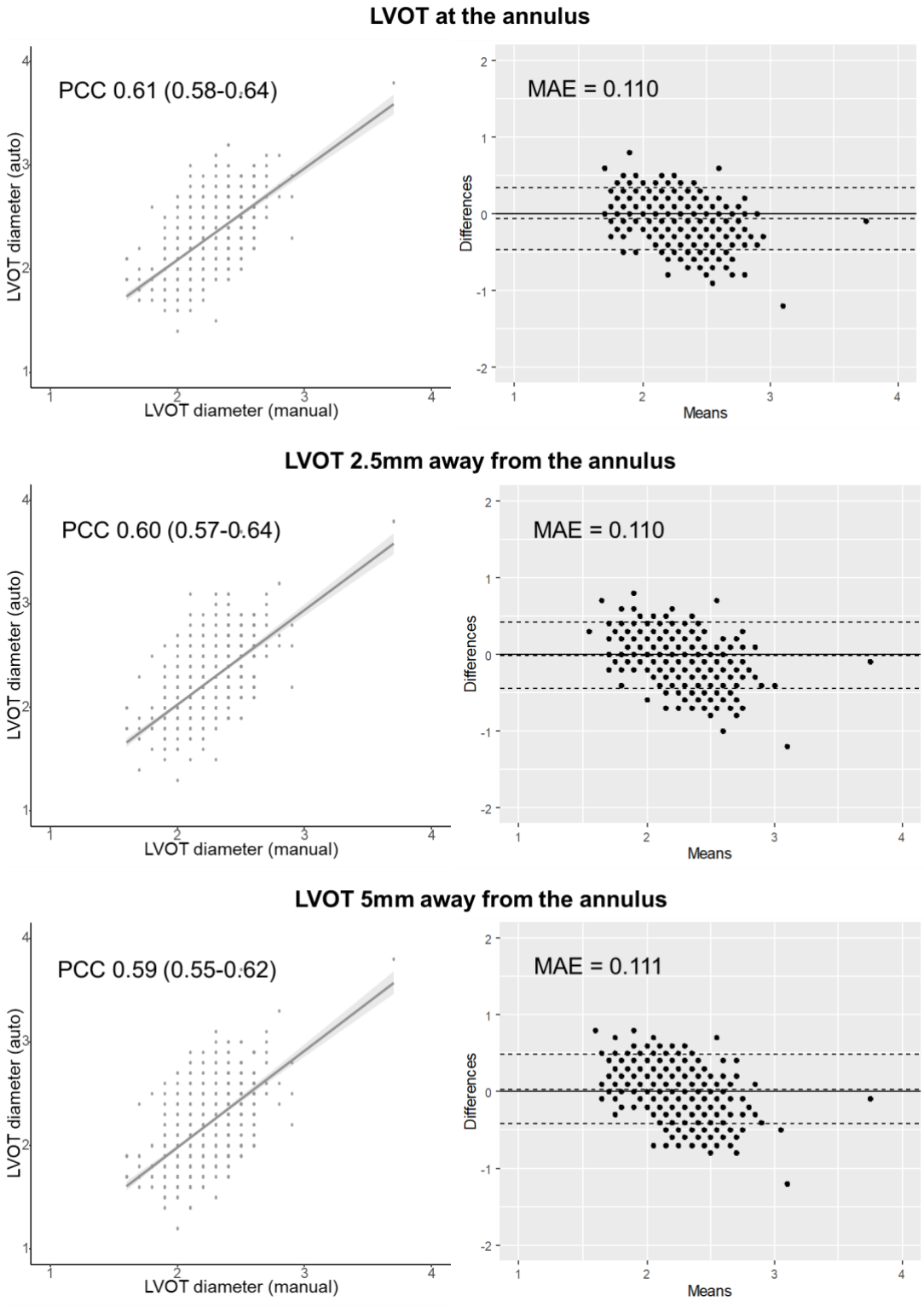
Supplemental Methods 6. Automatic Measurement of LVOT Diameter

From the predicted segmentation mask, we identify points where the MV intersects with the aorta and where the septum intersects with the aorta to determine annulus points. Using these points, we measure the LVOT diameter at three locations: 1) at the annulus, 2) 2.5mm away from the annulus towards the LV cavity, and 3) 5mm away from the annulus towards the LV cavity. This approach reflects differing opinions on the appropriate location for measuring the LVOT diameter.[10] Complete videos are presented in **Video S1**.

Automatic LVOT Measurements From the Predicted Segmentation Mask



For this study, the measurements taken at the annulus were used for analysis as they showed the highest agreement with the ground truth data.



Supplemental Methods 7. Availability of Ground Truth Measurements and Success Rate of Auto-Measurements in the AVS Group

The availability of ground truth measurements and the success rate of our algorithm's auto-measurements in the AVS group are as follows.

	Ground truth (% of the overall case)	**Auto-measurement (% of the overall case)**	**Matching case (% of available GT cases)**
ITDS (n=328)			
AV Vmax	328 (100)	328 (100)	328 (100)
AV mPG	320 (97.6)	328 (100)	320 (100)
LVOT VTI	166 (50.6)	164 (50.0)	164 (98.8)
LVOT diameter	159 (48.5)	290 (88.4)	141 (88.7)
AVA	156 (47.6)	143 (43.6)	133 (85.3)
DHDS (n=659)			
AV Vmax	83 (12.6)	652 (98.9)	83 (100)
AV mPG	602 (91.4)	652 (98.9)	598 (99.3)
LVOT VTI	438 (66.5)	583 (88.5)	367 (83.8)
LVOT diameter	425 (64.5)	618 (93.8)	405 (95.3)
AVA	560 (85.0)	543 (82.4)	457 (81.6)
TDDS (n=443)			
AV Vmax	443 (100)	443 (100)	443 (100)
AV mPG	440 (99.3)	443 (100)	440 (100)
LVOT VTI	235 (53.0)	233 (52.6)	233 (99.1)
LVOT diameter	228 (51.5)	419 (94.6)	212 (93.0)
AVA	227 (51.2)	219 (48.8)	209 (96.5)

AV, aortic valve; AVA, aortic valve area; LVOT, left ventricle outflow tract; mPG, mean pressure gradient; V_{max}, peak aortic valve velocity; VTI, velocity time integral.

Supplemental Results 1. Baseline Clinical Characteristics

Variables	ITDS	DHDS	TDDS
Entire study population, n	841	1,696	772
Age, years	69 (47-79)	57 (41-70)	77 (49-83)
Male, n (%)	424 (50.4)	842 (50.3)	390 (50.5)
Body mass index, kg/m^2	24 (22-26)	23 (21-25)	25 (22-27)
Patients with AVS, n	328	659	443
Hypertension, n (%)	87 (26.5)	412 (64)	96 (21.7)
Diabetes, n (%)	51 (15.5)	179 (27.8)	78 (17.6)

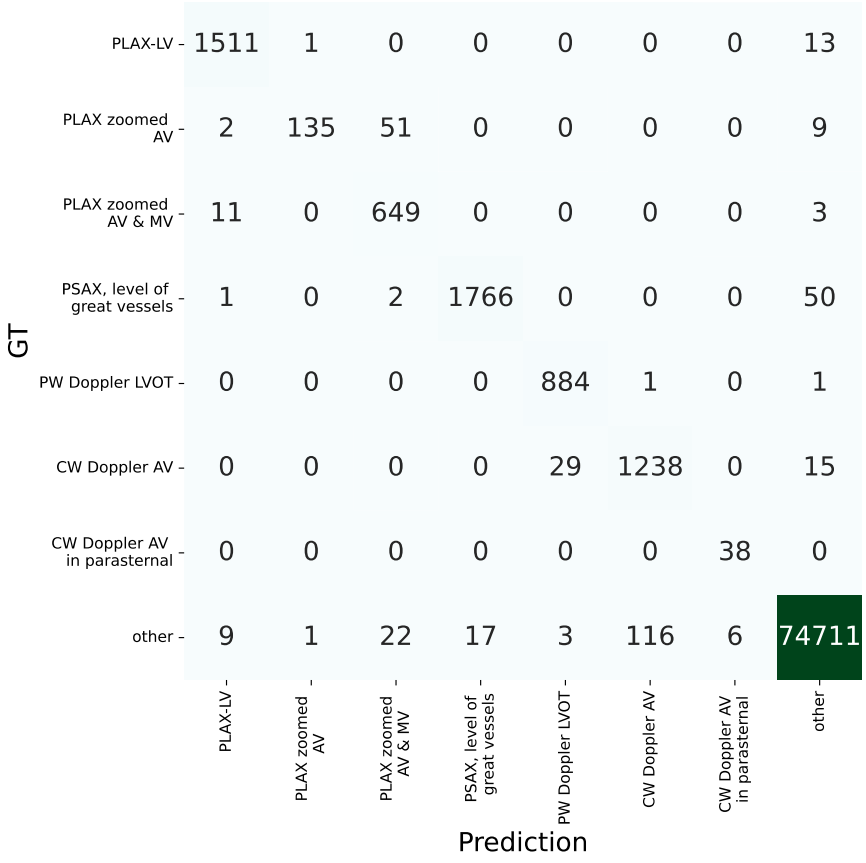
Values are given as numbers (percentage) or median (interquartile range)
Abbreviations: AVS, aortic valve stenosis; DHDS, Distinct Hospital Dataset; ITDS, Internal Test Dataset; TDDS, Temporally Distinct Dataset

Supplemental Results 2. View Classification Performance in Each Dataset

ITDS

	n	Precision	Recall	F1-score	Accuracy
PLAX	1,525	0.985	0.991	0.988	0.996
PLAX zoomed AV	197	0.985	0.685	0.808	
PLAX zoomed AV & MV	663	0.896	0.979	0.936	
PSAX at the level of AV	1,819	0.990	0.971	0.981	
PW doppler LVOT	886	0.965	0.998	0.981	
CW Doppler AV from apical views	1,282	0.914	0.966	0.939	
CW Doppler AV from the right parasternal view	38	0.864	1.000	0.927	
Other	74,885	0.999	0.998	0.998	

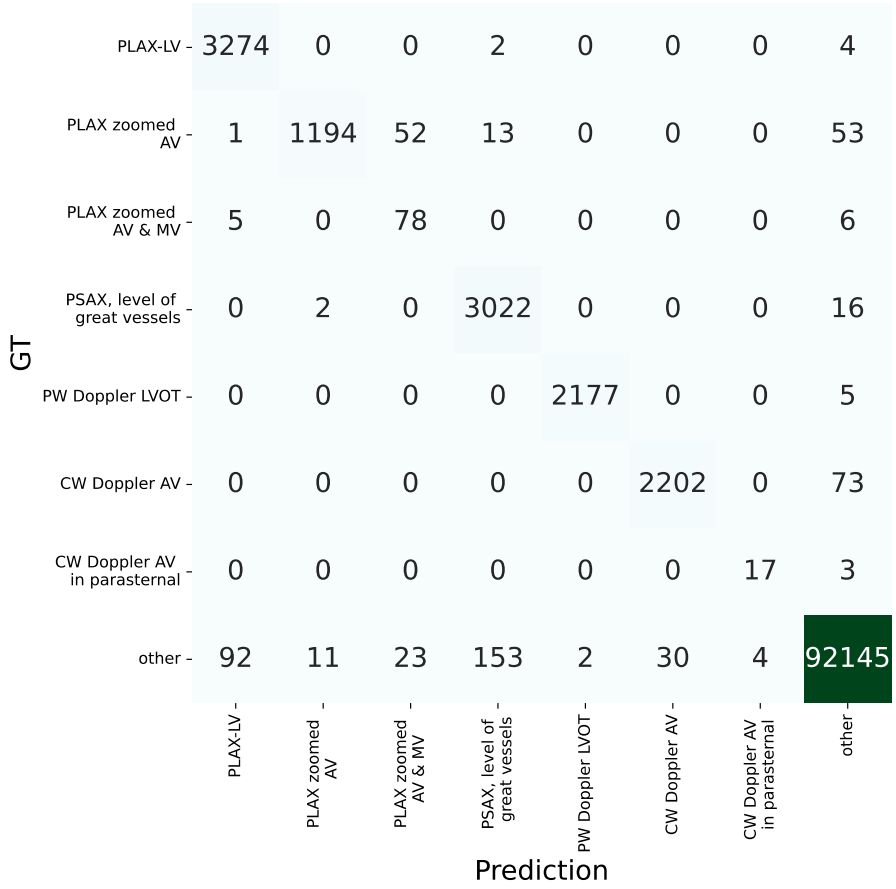
AV, aortic valve; CW, continuous wave Doppler; ITDS, internal test dataset; LVOT, left ventricle outflow tract; MV, mitral valve; PLAX, parasternal long-axis view. PSAX, parasternal short-axis view; PW, pulsed wave Doppler



DHDS

	n	Precision	Recall	F1-score	Accuracy
PLAX	3,280	0.971	0.998	0.984	0.995
PLAX zoomed AV	1,313	0.989	0.909	0.948	
PLAX zoomed AV & MV	89	0.510	0.876	0.645	
PSAX at the level of AV	3,040	0.947	0.994	0.970	
PW doppler LVOT	2,182	0.999	0.998	0.998	
CW Doppler AV from apical views	2,275	0.987	0.968	0.977	
CW Doppler AV from the right parasternal view	20	0.810	0.850	0.829	
Other	92,460	0.998	0.997	0.997	

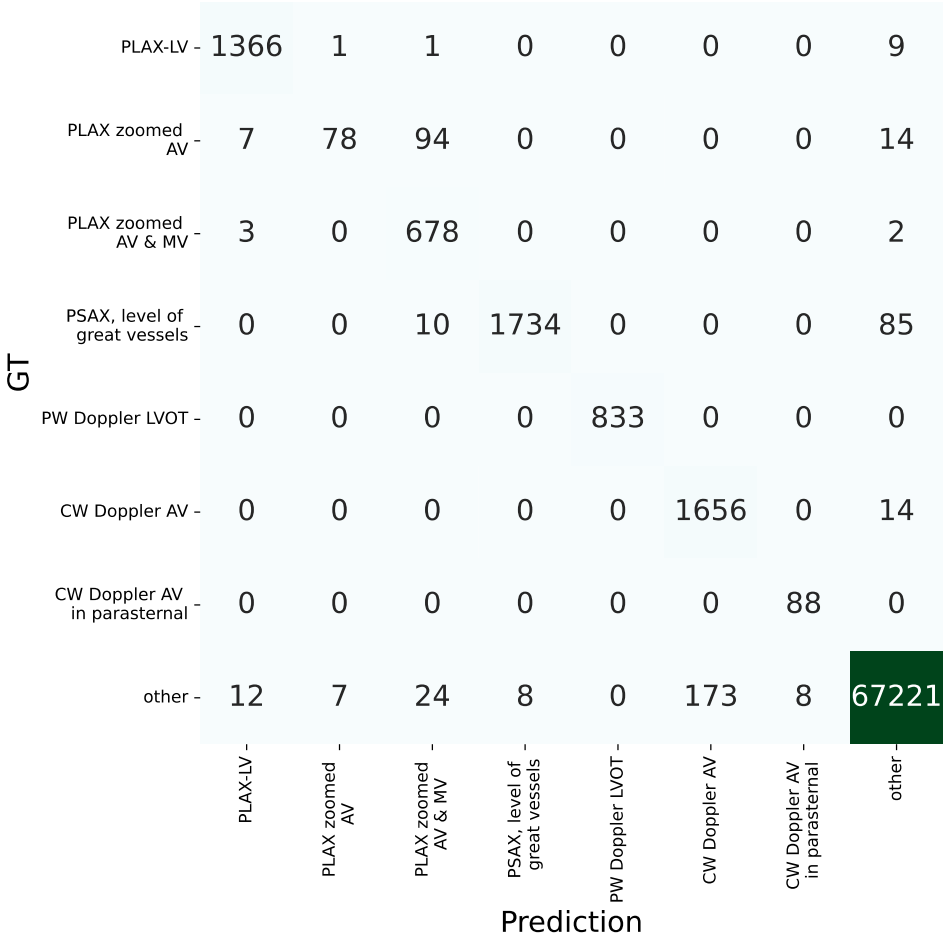
AV, aortic valve; CW, continuous wave Doppler; DHDS, Distinct Hospital Dataset; LVOT, left ventricle outflow tract; MV, mitral valve; PLAX, parasternal long-axis view. PSAX, parasternal short-axis view; PW, pulsed wave Doppler



TDDS

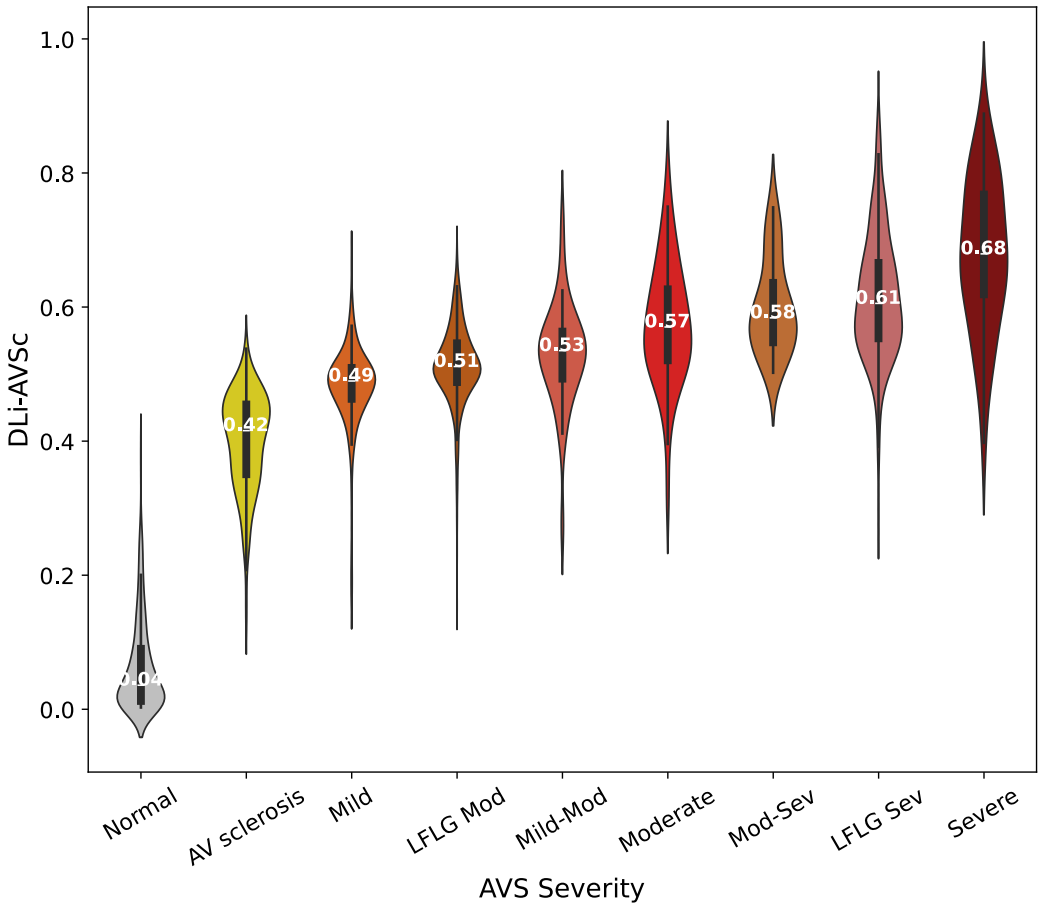
	n	Precision	Recall	F1-score	Accuracy
PLAX	1,377	0.984	0.992	0.988	
PLAX zoomed AV	193	0.907	0.404	0.559	
PLAX zoomed AV & MV	683	0.840	0.993	0.910	
PSAX at the level of AV	1,829	0.995	0.948	0.971	
PW doppler LVOT	833	1.000	1.000	1.000	0.994
CW Doppler AV from apical views	1,670	0.905	0.992	0.947	
CW Doppler AV from the right parasternal view	88	0.917	1.000	0.957	
Other	67,453	0.998	0.997	0.997	

AV, aortic valve; CW, continuous wave Doppler; TDDS, Temporally Distinct Dataset; LVOT, left ventricle outflow tract; MV, mitral valve; PLAX, parasternal long-axis view. PSAX, parasternal short-axis view; PW, pulsed wave Doppler



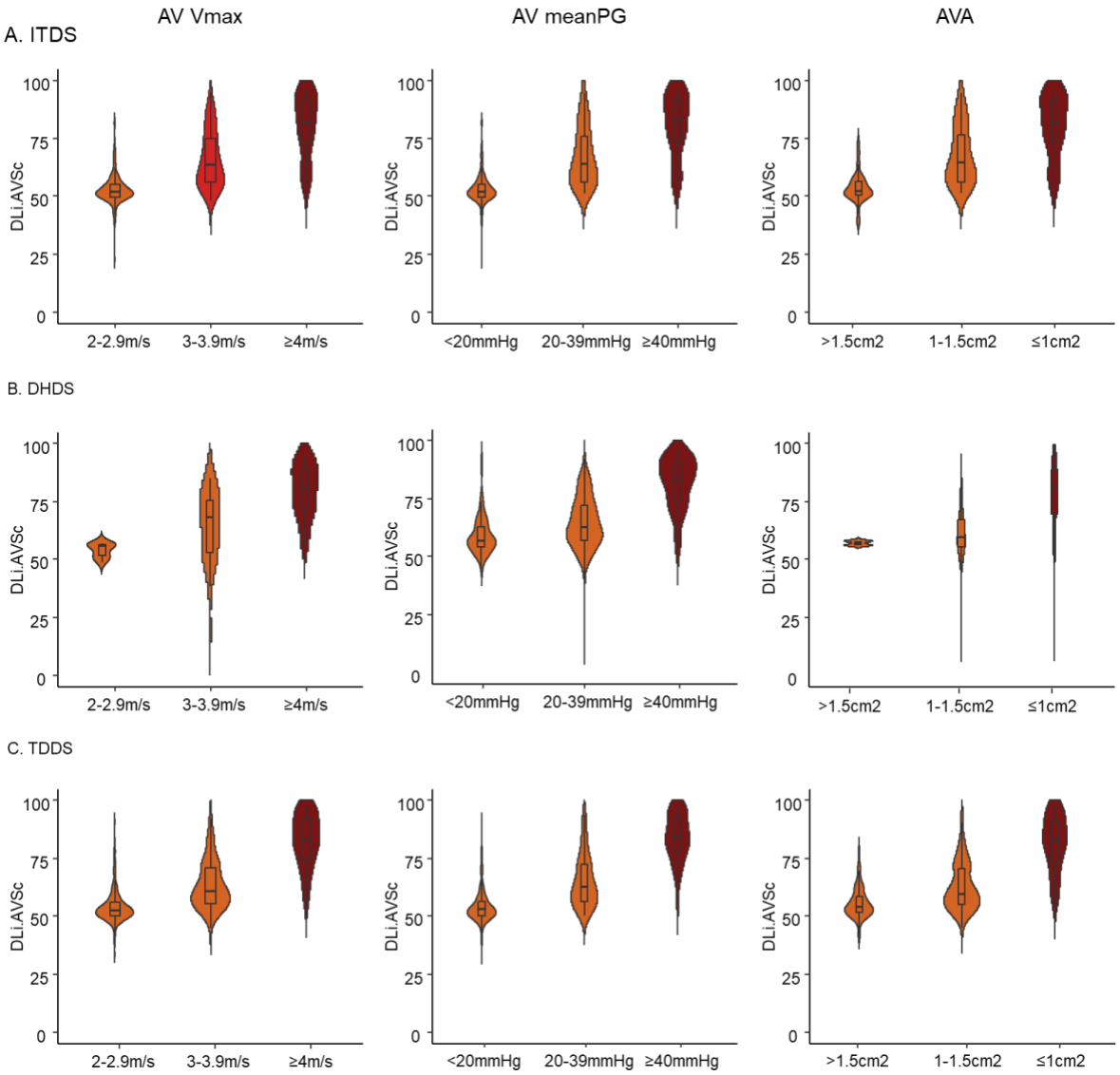
Supplemental Results 3. Distribution of DLi-AVSc in Discordant Cases of AVS Severity

In this study, discordant cases were defined as those where AV V_{max}, mPG, and AVA did not consistently fall into a single AVS severity class, resulting in interpretations spanning two classes, such as mild to moderate, moderate to severe, or cases of low-flow, low-gradient AVS where reduced stroke volume results in a lower pressure gradient despite significant stenosis. These cases were excluded from training, validation, and testing in the DDS. However, when included in the internal test dataset (ITDS) for comparison, the distribution of DLi-AVSc was as follows. This result suggests that DLi-AVSc can be helpful in cases where traditional parameters are discordant, making the assessment of AVS severity challenging.

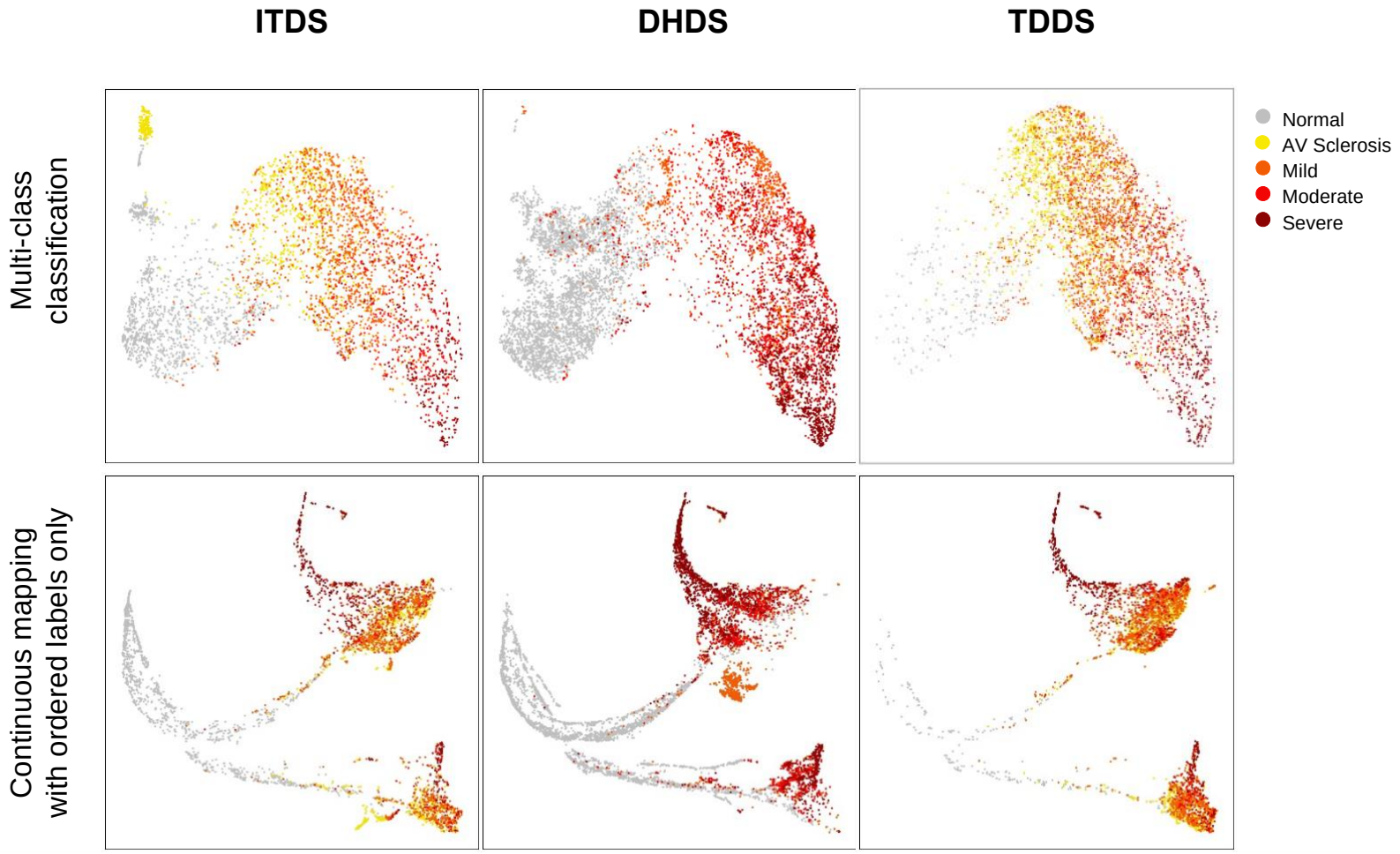


LFLG Mod, low-flow low-gradient moderate aortic stenosis; LFLG Sev, low-flow low-gradient severe aortic stenosis

Supplemental Results 4. Distribution of DLi-AVSc According to Conventional AVS Parameters

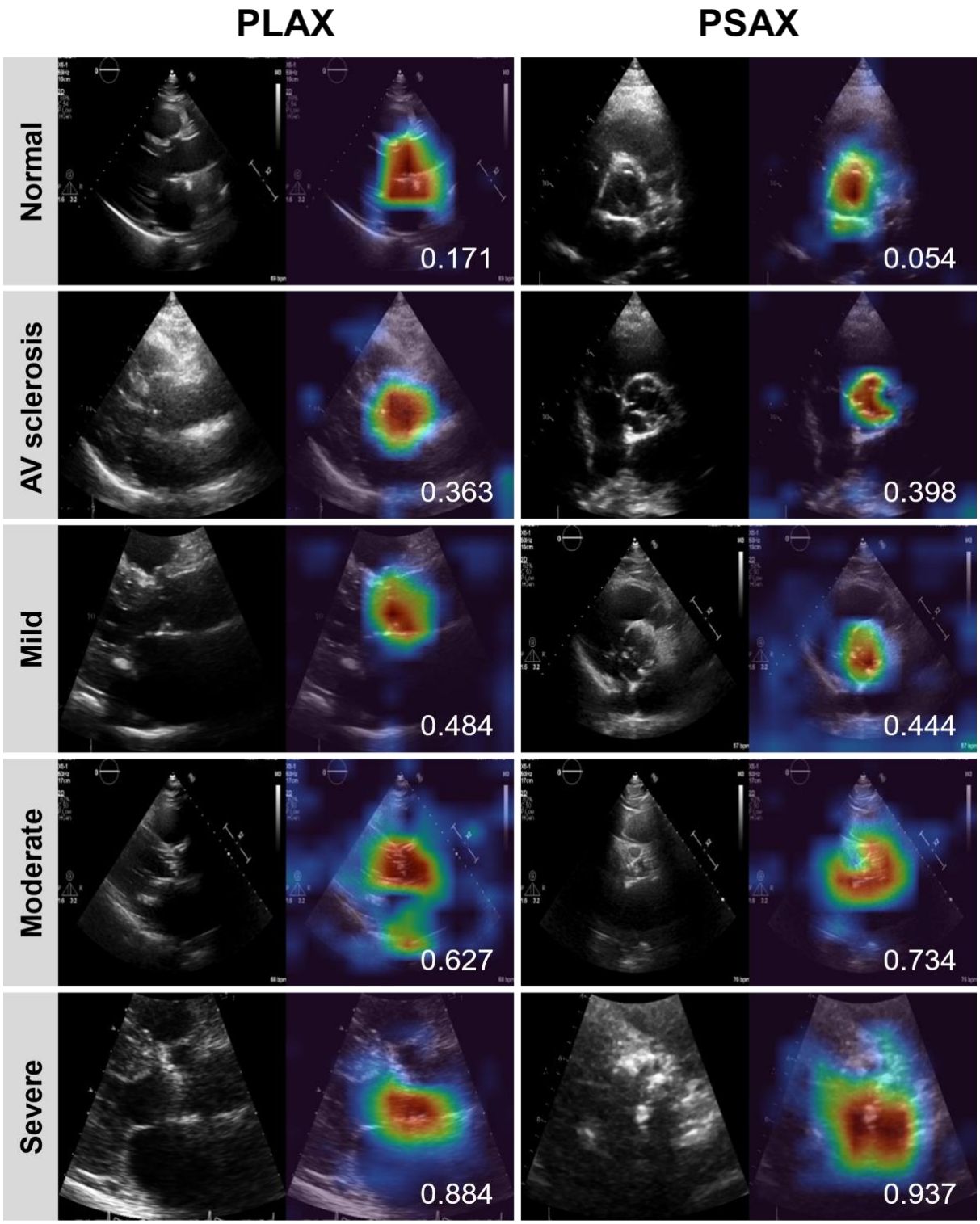


Supplemental Results 5. UMAP Visualization of AVS Continuum Using Different Approaches



In the first row, multi-class classification uses standard 5-class cross-entropy loss with one-hot encoding. In the second row, the network is trained using only continuous mapping with ordered labels, without multi-task learning with auxiliary tasks.

Supplemental Results 6. Explainability Analysis using Saliency Map



Given our use of video input with an r2plus1d network, we selected the frame with the maximum saliency value based on pixel-wise summation to provide a 2D saliency map image. Complete videos are presented in **Video S3.**

Reference

1. Jeon J, Ha S, Yoon Y, et al. Echocardiographic view classification with integrated out-of-distribution detection for enhanced automatic echocardiographic analysis. *arXiv preprint arXiv:2308.16483v1*. 2023.

2. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A Closer Look at Spatiotemporal Convolutions for Action Recognition. In: *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*. 2018:6450-6459.

3. Yu C, Gao C, Wang J, Yu G, Shen C, Sang N. BiSeNet V2: Bilateral Network with Guided Aggregation for Real-Time Semantic Segmentation. *International Journal of Computer Vision*. 2021;129:3051-3068.

4. Park J, Jeon J, Yoon YE, et al. Artificial intelligence-enhanced automation of left ventricular diastolic assessment: a pilot study for feasibility, diagnostic validation, and outcome prediction. *Cardiovasc Diagn Ther*. 2024;14:352-366.

5. Jeon J, Kim J, Jang Y, et al. A Unified Approach for Comprehensive Analysis of Various Spectral and Tissue Doppler Echocardiography. *arXiv preprint arXiv:2311.08439*. 2023.

6. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez JM, Luo P. SegFormer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*. 2021;34:12077-12090.

7. Cubuk ED, Zoph B, Shlens J, et al. Randaugment: Practical automated data augmentation with a reduced search space. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020:702-703.

8. Ostvik A, Salte IM, Smistad E, et al. Myocardial Function Imaging in Echocardiography Using Deep Learning. *IEEE Trans Med Imaging*. 2021;40:1340-1351.

9. Loshchilov I, Hutter F. Sgdr: Stochastic gradient descent with warm restarts. *arXiv*

preprint arXiv:160803983. 2016.

10. Baumgartner HC, Hung JC-C, Bermejo J, et al. Recommendations on the echocardiographic assessment of aortic valve stenosis: a focused update from the European Association of Cardiovascular Imaging and the American Society of Echocardiography. *Eur Heart J Cardiovasc Imaging*. 2017;18:254-275.