This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

S1 Supplementary Results

S1.1 EchoFM Intermediate Representations

To address the potential benefits of utilizing intermediate representations from Masked Autoencoders (MAEs), we evaluated the downstream performance of EchoFM across the latter half of its 24-layer transformer stack. As detailed in Table S1, extracting features from Layer 16 yields a slight improvement in the mean F1-score (0.356±0.052) compared to the final layer (0.341±0.039). However, the final layer maintains the highest overall ROC-AUC (0.628±0.039). Because these performance differences are marginal and fall well within the margins of standard deviation, we conclude that intermediate layer extraction does not provide a substantial enough advantage to warrant deviating from our standardized, last-layer evaluation protocol.

Table S1: Performance metrics for 4-class mitral regurgitation classification ablated for intermediate representations of the 24-layer EchoFM vision transformer. Parameters represent probe only. Best results per metric are in **bold**, second best underlined.

	Model	Params	F1	ROC-AUC	Specificity	Sensitivity	Precision	NPV
Attentive	Layer 14	11.5M	.339 ± .044	.621 ± .036	.776 ± .018	.387 ± .047	.319 ± .046	.796 ± .020
	Layer 16	11.5M	**.356** ± .052	.625 ± .038	**.785** ± .020	**.409** ± .055	**.336** ± .055	**.807** ± .023
	Layer 18	11.5M	.346 ± .055	.622 ± .039	.780 ± .024	.397 ± .062	.325 ± .057	.801 ± .028
	Layer 20	11.5M	.348 ± .045	.622 ± .038	.781 ± .019	.400 ± .051	.326 ± .043	.802 ± .023
	Last Layer	11.5M	.341 ± .039	**.628** ± .039	.775 ± .017	.387 ± .044	.319 ± .038	.795 ± .020