

High Throughput Deep Learning Detection of Mitral Regurgitation

Amey Vrudhula B.S.E.^{a,b}, Grant Duffy B.S.^a, Milos Vukadinovic B.S.^{a,c}, David Liang M.D.,
Ph.D.^d, Susan Cheng, M.D., M.M.Sc., M.P.H.^a, David Ouyang M.D.^{a,e}

a) Department of Cardiology, Smidt Heart Institute, Cedars-Sinai Medical Center, Los Angeles, CA

b) Icahn School of Medicine at Mt Sinai, New York, NY

c) Department of Bioengineering, University of California Los Angeles, Los Angeles, CA

d) Department of Medicine, Division of Cardiology, Stanford University, Palo Alto, CA

e) Division of Artificial Intelligence in Medicine, Cedars-Sinai Medical Center, Los Angeles, CA

Short Title: Deep Learning Detection of Mitral Regurgitation

Total Word Count: 2,423

Correspondence:

David Ouyang M.D.

Assistant Professor

Smidt Heart Institute

Cedars-Sinai Medical Center

127 S. San Vicente Blvd.

AHSP Pavilion Suite A3100,

Los Angeles, CA 90048

Telephone: [310-423-3475](tel:310-423-3475)

Fax: [310.423.0127](tel:310.423.0127)

Email: David.Ouyang@cshs.org

Twitter Handle: [@David_Ouyang](https://twitter.com/David_Ouyang)

32 **Abstract**

33 **Background:** Diagnosis of mitral regurgitation (MR) requires careful evaluation of
34 echocardiography with Doppler imaging. This study presents the development and validation of a
35 fully automated deep learning pipeline for identifying apical-4-chamber view videos with color
36 Doppler and detection of clinically significant (moderate or severe) mitral regurgitation from
37 transthoracic echocardiography studies.

38 **Methods:** A total of 58,614 studies (2,587,538 videos) from Cedars-Sinai Medical Center (CSMC)
39 were used to develop and test an automated pipeline to identify apical-4-chamber view videos with
40 color Doppler across the mitral valve and then assess mitral valve regurgitation severity. The
41 model was tested on an internal test set of 1,800 studies (80,833 videos) from CSMC and externally
42 evaluated in a geographically distinct cohort of 915 studies (46,890 videos) from Stanford
43 Healthcare (SHC).

44 **Results:** In the held-out CSMC test set, the view classifier demonstrated an AUC of 0.998 (0.998
45 - 0.999) and correctly identified 3,452 of 3,539 MR color Doppler videos (sensitivity of 0.975
46 (0.968-0.982) and specificity of 0.999 (0.999-0.999) compared with manually curated videos). In
47 the external test cohort from SHC, the view classifier correctly identified 1,051 of 1,055 MR color
48 Doppler videos (sensitivity of 0.996 (0.990 – 1.000) and specificity of 0.999 (0.999 – 0.999)
49 compared with manually curated videos). For evaluating clinically significant MR, in the CSMC
50 test cohort, moderate-or-severe MR was detected with AUC of 0.916 (0.899 - 0.932) and severe
51 MR was detected with an AUC of 0.934 (0.913 - 0.953). In the SHC test cohort, the model detected
52 moderate-or-severe MR with an AUC of 0.951 (0.924 - 0.973) and severe MR with an AUC of
53 0.969 (0.946 - 0.987).

54 **Conclusions:** In this study, we developed and validated an automated pipeline for identifying
55 clinically significant MR from transthoracic echocardiography studies. Such an approach has
56 potential for automated screening of MR and precision evaluation for surveillance.

57
58
59
60
61
62

63 **Introduction**

64 Mitral regurgitation (MR) is one of the most common forms of valve disease, affecting
65 more than 4 million Americans.^{1,2,3} Often progressing insidiously and frequently
66 underrecognized³, both primary MR as well as secondary MR can be initially asymptomatic but
67 lead to worsening heart failure and mortality.^{1,4-6,7,8} There has been an increased focus on early
68 MR diagnosis given advances in surgical and transcatheter treatment options^{1,9,10}.
69 Echocardiography with color Doppler is the most common method of initial evaluation of MR,
70 with a holistic assessment combining left atrial size, effective regurgitant orifice area, regurgitant
71 fraction, regurgitant volume, as well as other key clinical factors to accurately assess disease
72 severity.^{11,12} Despite ultrasound technology becoming more widely available, accurate assessment
73 of MR still requires experienced expert image acquisition and evaluation.

74 Recent advances in machine learning offer opportunities to automate time-consuming steps
75 in the interpretation of medical imaging. Artificial intelligence (AI) has the ability to precisely
76 phenotype subtle cardiac physiology as well as identify imaging features of disease severity not
77 recognized by clinicians.¹³⁻¹⁶ Deep learning has been applied to echocardiography to improve the
78 precision of common measurements, such as left ventricular ejection fraction¹³ and wall
79 thickness^{15,17}, as well as streamlining assessment of aortic stenosis¹⁸, hypertrophic
80 cardiomyopathy (HCM)¹⁵, and cardiac amyloidosis (CA).^{15,19} With increased ultrasound
81 availability, AI guidance has been developed for both image acquisition and interpretation^{13,20}.
82 With the increasing prevalence of MR in an aging population with co-morbid heart failure, AI
83 could aid in MR screening and surveillance.²¹⁻²⁵

84 In this study, we developed and evaluated a deep learning pipeline's performance in
85 automating identification of MR from standard transthoracic echocardiogram studies. We
86 hypothesized that a deep learning approach can identify color Doppler apical-4-chamber videos
87 and assess MR severity with high-throughput automation, and this automated pipeline was
88 evaluated in two geographically distinct cohorts (**Figure 1**). Combined with other
89 echocardiography AI algorithms, such an approach can be used for serial surveillance and
90 screening of mitral regurgitation.

91
92
93

94 **Methods**

95 **Study Population and Data Source**

96 ***Cedars-Sinai Medical Center (CSMC) Cohort:*** A total of 58,614 transthoracic
97 echocardiogram studies from 38,461 patients receiving care at Cedars-Sinai Medical Center
98 (CSMC) between October 11, 2011 and June 04, 2022 were used to train and evaluate the deep
99 learning models. A total of 2,587,538 videos (an average of 44 videos per study after excluding
100 still images) were initially sourced from Digital Imaging and Communications in Medicine
101 (DICOM) files and underwent de-identification, view classification, and pre-processing into AVI
102 videos. 354,117 videos were classified as apical-4-chamber videos using an automated view
103 classifier, and then manually curated to identify 34,714 videos with color Doppler across the mitral
104 valve.²⁶

105 Following view selection, the CSMC cohort included 34,714 videos across 30,453 unique
106 echocardiogram studies from 22,661 patients, and a subset enriched for moderate and severe MR
107 were used for training. A total of 20,604 videos from 18,133 studies in the dataset were split on a
108 patient level into train (80%), validation (10%), and test (10%) cohorts to train a deep neural
109 network for MR severity classification (**Figure 2**). MR severity for each study was determined
110 based on the clinical echocardiogram report determined in a high volume echocardiography lab in
111 accordance with ASE guidelines.²⁷ When MR was characterized as an intermediate category
112 (“trace to mild” or “mild to moderate” or “moderate to severe”), videos were placed in the more
113 severe categories. Both primary and secondary MR were included. Studies with concomitant mitral
114 stenosis, other prosthetic valves, and heart failure were also included in both training and
115 validation datasets.

116 ***Stanford Healthcare (SHC) Cohort:*** The pipeline was evaluated on 915 studies
117 (containing a total of 46,890 videos) from SHC’s high-volume academic echocardiography lab.
118 The automated view classification pipeline was compared with manual curation of videos within
119 those studies to evaluate specificity. All videos identified by the view classifier were used for
120 downstream MR severity model validation. Model output was compared with MR severity
121 determined by expert cardiologists from the clinical reports. This study was approved by the
122 Institutional Review Boards at Cedars-Sinai Medical Center and Stanford Healthcare. The need
123 for informed consent was waived as the study involved secondary analysis of existing data.

124

125 **AI Model Training**

126 Deep learning models were trained using the PyTorch Lightning deep learning framework.
127 When patients had multiple echocardiogram videos and studies, each video was considered an
128 independent example during training, with care not to have patient overlap across training,
129 validation, and test cohorts. Video-based convolutional neural networks (R2+1D) were used for
130 view classification and MR severity assessment.²⁸ This model architecture was previously used
131 for other echocardiography tasks and shown to be effective.¹⁷ The models were initialized with
132 random weights and trained using a binary cross entropy loss function for up to 100 epochs, using
133 an ADAM optimizer, an initial learning rate of 1e-2, and a batch size of 24 on two NVIDIA RTX
134 3090 GPUs. Early stopping was performed based on the validation loss.

135 The view classifier was trained using the 34,714 manually curated videos with color
136 Doppler across the mitral valve as cases and 49,263 other apical-4-chamber videos as controls.
137 Controls were a combination of videos that did not have color Doppler or had color Doppler
138 window not focused on the mitral valve (videos focused on the tricuspid valve, intra-atrial septum,
139 or ventricular septum). The MR severity model was trained on 6,206 videos without MR, 6,128
140 videos with mild MR, 6,174 videos with moderate MR and 2,042 videos with severe MR. This
141 process is summarized in **Figure 2**.

142

143 **Statistical Analysis**

144 Model performance was evaluated using area under the receiver operating characteristic
145 curve (AUROC) and confusion matrices. F1-score, recall (sensitivity), positive predictive value
146 (PPV), and negative predictive value (NPV) for both greater than moderate MR and severe MR.
147 During external validation, the view classifier and MR classifier were evaluated serially as an
148 automated pipeline. Statistical analysis was performed in Python (version 3.8.0) and R (version
149 4.2.2). Confidence intervals were computed via bootstrapping with 10,000 samples. Reporting of
150 study results are consistent with guidelines put forth by CONSORT-AI.^{29,30}

151

152 **Model Explainability**

153 The key imaging features identified by the MR severity model were evaluated using
154 saliency mapping generated using the Integrated Gradients method.³¹ This method generated a
155 heatmap for every frame of the video, summarized as a final 2-dimensional heatmap generated by

156 using the maximum value along the temporal axis for each pixel location in the video. Pixels
157 brighter in intensity and closer to yellow were more salient to model predictions, while those darker
158 in color were less important to the model's final prediction. When assessing videos with no MR,
159 heatmaps were obtained by taking the maximum of saliency maps for the moderate and severe
160 class output neurons for each pixel location (**Figure 4**).

161

162

163

164

165

166

167

168

169

170

171

172

173

174

175

176

177

178

179

180

181

182

183

184

185

186

187

188 **Results:**

189 **Study Population**

190 A total of 58,614 studies from 38,461 patients were used to train the deep learning pipeline.
191 From 2,587,538 initial videos, a total of 354,117 videos were identified as apical-4-chamber and
192 subsequently manually curated to identify videos that had color Doppler across the mitral valve.
193 The manually curated color Doppler videos were used to train a view-classification model and
194 linked with clinician reports to train the MR severity model. Patient characteristics are presented
195 in **Tables 1 & 2** and are representative of the general CSMC patient population that received
196 echocardiograms. The data was split on patient level for training and validation and had similar
197 patient age, ejection fraction, left atrial volume index and proportions of male sex, coronary artery
198 disease, and atrial fibrillation.

199

200 **View Classifier Performance Across Two Institutions**

201 On a test set of 3,109 studies (132,767 videos) from CSMC not seen during model training,
202 the view classifier identified 3,452 videos (97.5% of manually identified cases). This corresponds
203 to an of AUC of 0.998 (0.998 – 0.999), and at the Youden Index, with a sensitivity of 0.975 (0.968-
204 0.982) and specificity of 0.999 (0.999-0.999). To evaluate generalization of the view classification
205 model at a geographically distinct site, we evaluated its performance on 915 studies from SHC.
206 The view classifier isolated 1,091 videos from a total of 46,890 videos, while manual review
207 identified 1,055 videos with color Doppler across the mitral valve. The view classifier correctly
208 identified 1,051 (99.6%) of manually curated videos, with 4 videos not found by the AI pipeline
209 and 40 false positives. This corresponds to a sensitivity of 0.996 (0.990 – 1.000) and specificity of
210 0.999 (0.999 – 0.999).

211

212 **Mitral Regurgitation Severity Performance Across Two Institutions**

213 The MR severity model showed strong performance in distinguishing MR severity and
214 identifying clinically significant mitral regurgitation (**Figure 3**). In the internal CSMC test set not
215 used during model training, the model demonstrated an AUC of 0.916 (0.899 - 0.932) in detecting
216 \geq moderate MR and an AUC of 0.934 (0.913 - 0.953) for severe MR. The AI model had an NPV
217 of 0.954 (0.940 - 0.967) for severe MR and an NPV of 0.863 (0.835 - 0.890) for \geq moderate MR.

218 Further information on MR model performance is presented in **Table 3**. The MR severity model
219 performance was similar across institutions. In the SHC cohort, the model identified severe MR
220 with an AUC of 0.969 (0.946 - 0.987) and \geq moderate MR with an AUC of 0.951 (0.924 - 0.973).
221 In this cohort, the model had an NPV of 0.977 (0.962 – 0.990) for severe MR and an NPV of 0.986
222 (0.974 – 0.995) for \geq moderate MR.

223

224 **Model Interpretation**

225 Notably, saliency maps for our model demonstrate that the model focuses on the clinically
226 relevant imaging features of mitral regurgitation. Saliency maps from Integrated Gradients were
227 used to identify regions of interest in each video contributing the most to detection of MR severity
228 (**Figure 4**).³¹ These interpretability techniques demonstrated localization of the activation signal
229 in the color Doppler window and primarily highlighting the regurgitant jet, indicating that the
230 model used appropriate, physiologic features of the mitral regurgitation to make predictions.
231 Frame-by-frame saliency visualizations are shown in **Supplemental Videos S1-S4**.

232

233

234

235

236

237

238

239

240

241

242

243

244

245

246

247

248

249

250 **Discussion**

251 We developed and validated an automated pipeline for assessing for mitral regurgitation in
252 echocardiography. From a full transthoracic echocardiogram study, the algorithm automatically
253 screens for appropriate A4C videos with color Doppler on the mitral valve and then assesses MR
254 severity. For both severe MR as well as \geq moderate MR, the model demonstrated strong
255 performance (> 0.916 AUC and > 0.863 NPV). This automated workflow worked in unselected
256 external validation studies without preselection or exclusion of other concomitant comorbidities.
257 Given these characteristics, our deep learning model could aid in the preliminary assessment of
258 MR, facilitate review of institutional databases, or expand access for screening in low-resource
259 settings.

260 Our algorithm learns features of mitral regurgitation that generalize across variability in
261 imaging practices in two geographically distinct sites. Many prior echocardiography AI models
262 primarily focused on black-and-white standard 2D B-mode images, while our study focuses on the
263 AI assessment of color Doppler videos and utilized a video-based model for the incorporation of
264 rich temporal information, both crucial for accurate MR assessment. Incorporation of Doppler
265 information greatly expands the opportunities for AI in echo, particularly in valve disease. In
266 expert clinical interpretation, a variety of metrics beyond just color Doppler and views are
267 synthesized together to come up with a holistic assessment of MR severity. Intriguingly, our AI
268 algorithm generally results in concordant interpretations with the comprehensive clinical approach
269 while relying only on the A4C view, suggesting there is significant overlapping information as
270 well as dependence on the A4C view in standard clinical practice.

271 While promising, the present work carries limitations. Echocardiographic assessment of
272 MR depends on appropriate images being obtained, with different views potentially maximizing
273 the visualized regurgitant jet. This algorithm would not overcome incomplete input information
274 and insufficient images that would result underestimation of MR. Future work could focus on
275 automatically quantifying parameters like valve leaflet thickness, effective regurgitant orifice area,
276 regurgitant volume and fraction.

277 The present work builds upon prior work in the space of echocardiography and AI. Several
278 recent works have reported strides in computer vision and echocardiography, including automated
279 view classification^{32,33}, phenotyping of left ventricular hypertrophy¹⁵, assessment of LV systolic

280 function¹³, aortic stenosis risk stratification, and detection of complex congenital heart defects.³⁴
281 Prior work in machine learning applied to MR has primarily focused on structured data and non-
282 deep learning approaches. The combination of our algorithm with previously published works
283 using AI to guide novices in acquiring imaging could potentially increase access to screening of
284 MR.^{20,35}

285 In summary, we introduce a model to screen for and stratify mitral regurgitation severity
286 from transthoracic echocardiogram videos. To do so, we provide a workflow for isolating mitral
287 valve color Doppler videos and automation of MR severity assessment. The models were evaluated
288 to have good performance in internal and external test cohorts. The use of such a model, with a
289 high AUC, NPV, and generalizability across sites, can open the door for screening of mitral valve
290 disease in the primary care setting or in low-resource environments.

291
292 **Acknowledgements:** A.V. is a research fellow supported by the Sarnoff Cardiovascular Research
293 Award. S.C. acknowledges support from the Erika J Glazer Family Foundation.

294
295 **Disclosures:** A.V., G.D., M.V., and D.L. report no disclosures. S.C. reports consulting fees from
296 UCB and Viz.ai and research grants NIH R01-HL131532 and NIH R01-HL142983. D.O. reports
297 support from NIH NHLBI, NIH grant R00-HL157421 and AstraZeneca Alexion, as well as
298 consulting from EchoIQ, Ultromics, Pfizer, InVision, Korean Society of Echo, and Japanese
299 Society of Echo.

300
301 **Code Statement:** Our code and model weights are available at <https://github.com/echonet/MR>

302

303

References

- 304 1. Enriquez-Sarano M, Akins CW, Vahanian A. Mitral regurgitation. *Lancet*. 2009;373:1382–
305 1394.
- 306 2. Harb SC, Griffin BP. Mitral Valve Disease: a Comprehensive Review. *Curr Cardiol Rep*.
307 2017;19:73.
- 308 3. Dziadzko V, Clavel M-A, Dziadzko M, Medina-Inojosa JR, Michelena H, Maalouf J,
309 Nkomo V, Thapa P, Enriquez-Sarano M. Outcome and undertreatment of mitral
310 regurgitation: a community cohort study. *Lancet*. 2018;391:960–969.
- 311 4. Otto CM, Verrier ED. Mitral regurgitation--what is best for my patient? *N. Engl. J. Med*.
312 2011;364:1462–1463.
- 313 5. Del Forno B, De Bonis M, Agricola E, Melillo F, Schiavi D, Castiglioni A, Montorfano M,
314 Alfieri O. Mitral valve regurgitation: a disease with a wide spectrum of therapeutic options.
315 *Nat Rev Cardiol*. 2020;17:807–827.
- 316 6. Ocher R, May M, Labin J, Shah J, Horwich T, Watson KE, Yang EH, Calfon Press,
317 Marcella A. Mitral Regurgitation in Female Patients: Sex Differences and Disparities.
318 *Catheter Cardiovasc Interv*. 2023;2:101032.
- 319 7. Simpson TF, Kumar K, Samhan A, Khan O, Khan K, Strehler K, Fishbein S, Wagner L,
320 Sotelo M, Chadderdon S, Golwala H, Zahr F. Clinical Predictors of Mortality in Patients
321 with Moderate to Severe Mitral Regurgitation. *Am J Med*. 2022;135:380-385.e3.
- 322 8. Messika-Zeitoun D, Candolfi P, Vahanian A, Chan V, Burwash IG, Philippon J-F, Toussaint
323 J-M, Verta P, Feldman TE, Iung B, Glineur D, Mesana T, Enriquez-Sarano M. Dismal
324 Outcomes and High Societal Burden of Mitral Valve Regurgitation in France in the Recent
325 Era: A Nationwide Perspective. *J Am Heart Assoc*. 2020;9:e016086.
- 326 9. Tribouilloy CM, Enriquez-Sarano M, Schaff HV, Orszulak TA, Bailey KR, Tajik AJ, Frye
327 RL. Impact of preoperative symptoms on survival after surgical correction of organic mitral
328 regurgitation: rationale for optimizing surgical indications. *Circulation*. 1999;99:400–405.

- 329 10. David TE, Ivanov J, Armstrong S, Rakowski H. Late outcomes of mitral valve repair for
330 floppy valves: Implications for asymptomatic patients. *J Thorac Cardiovasc Surg.*
331 2003;125:1143–1152.
- 332 11. Fadel BM, Bakarman H, Dahdouh Z, Di Salvo G, Mohty D. Spectral Doppler interrogation
333 of mitral regurgitation -spot diagnosis. *Echocardiography.* 2015;32:1179–1183.
- 334 12. Hagendorff A, Knebel F, Helfen A, Stöbe S, Haghi D, Ruf T, Lavall D, Knierim J, Altiok E,
335 Brandt R, Merke N, Ewen S. Echocardiographic assessment of mitral regurgitation:
336 discussion of practical and methodologic aspects of severity quantification to improve
337 diagnostic conclusiveness. *Clin Res Cardiol.* 2021;110:1704–1733.
- 338 13. Ouyang D, He B, Ghorbani A, Yuan N, Ebinger J, Langlotz CP, Heidenreich PA,
339 Harrington RA, Liang DH, Ashley EA, Zou JY. Video-based AI for beat-to-beat assessment
340 of cardiac function. *Nature.* 2020;580:252–256.
- 341 14. Elias P, Poterucha TJ, Rajaram V, Moller LM, Rodriguez V, Bhave S, Hahn RT, Tison G,
342 Abreau SA, Barrios J, Torres JN, Hughes JW, Perez MV, Finer J, Kodali S, Khalique O,
343 Hamid N, Schwartz A, Homma S, Kumaraiah D, Cohen DJ, Maurer MS, Einstein AJ, Nazif
344 T, Leon MB, Perotte AJ. Deep Learning Electrocardiographic Analysis for Detection of
345 Left-Sided Valvular Heart Disease. *J Am Coll Cardiol.* 2022;80:613–626.
- 346 15. Duffy G, Cheng PP, Yuan N, He B, Kwan AC, Shun-Shin MJ, Alexander KM, Ebinger J,
347 Lungren MP, Rader F, Liang DH, Schnittger I, Ashley EA, Zou JY, Patel J, Witteles R,
348 Cheng S, Ouyang D. High-Throughput Precision Phenotyping of Left Ventricular
349 Hypertrophy With Cardiovascular Deep Learning. *JAMA Cardiol.* 2022;7:386–395.
- 350 16. He B, Kwan AC, Cho JH, Yuan N, Pollick C, Shiota T, Ebinger J, Bello NA, Wei J, Josan
351 K, Duffy G, Jujjavarapu M, Siegel R, Cheng S, Zou JY, Ouyang D. Blinded, randomized
352 trial of sonographer versus AI cardiac function assessment. *Nature.* 2023;616:520–524.
- 353 17. Soto JT, Weston Hughes J, Sanchez PA, Perez M, Ouyang D, Ashley EA. Multimodal deep
354 learning enhances diagnostic precision in left ventricular hypertrophy. *Eur Heart J Digit*
355 *Health.* 2022;3:380–389.

- 356 18. Holste G, Oikonomou EK, Mortazavi BJ, Coppi A, Faridi KF, Miller EJ, Forrest JK,
357 McNamara RL, Ohno-Machado L, Yuan N, Gupta A, Ouyang D, Krumholz HM, Wang Z,
358 Khera R. Severe aortic stenosis detection by deep learning applied to echocardiography. *Eur*
359 *Heart J* [Internet]. 2023; Available from: <http://dx.doi.org/10.1093/eurheartj/ehad456>
- 360 19. Vrudhula A, Stern L, Cheng PC, Ricchiuto P, Daluwatte C, Witteles R, Patel J, Ouyang D.
361 Impact of case and control selection on training AI screening of cardiac amyloidosis
362 [Internet]. bioRxiv. 2023; Available from: <http://dx.doi.org/10.1101/2023.03.30.23287941>
- 363 20. Narang A, Bae R, Hong H, Thomas Y, Surette S, Cadieu C, Chaudhry A, Martin RP,
364 McCarthy PM, Rubenson DS, Goldstein S, Little SH, Lang RM, Weissman NJ, Thomas JD.
365 Utility of a Deep-Learning Algorithm to Guide Novices to Acquire Echocardiograms for
366 Limited Diagnostic Use. *JAMA Cardiol.* 2021;6:624–632.
- 367 21. Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden
368 of valvular heart diseases: a population-based study. *Lancet.* 2006;368:1005–1011.
- 369 22. Martin A-C, Bories M-C, Tence N, Baudinaud P, Pechmajou L, Puscas T, Marijon E,
370 Achouh P, Karam N. Epidemiology, Pathophysiology, and Management of Native
371 Atrioventricular Valve Regurgitation in Heart Failure Patients. *Front Cardiovasc Med.*
372 2021;8:713658.
- 373 23. Avierinos J-F, Tribouilloy C, Grigioni F, Suri R, Barbieri A, Michelena HI, Ionico T,
374 Rusinaru D, Ansaldi S, Habib G, Szymanski C, Giorgi R, Mahoney DW, Enriquez-Sarano
375 M, Mitral regurgitation International DAtabase (MIDA) Investigators. Impact of ageing on
376 presentation and outcome of mitral regurgitation due to flail leaflet: a multicentre
377 international study. *Eur Heart J.* 2013;34:2600–2609.
- 378 24. Grave C, Tribouilloy C, Tuppin P, Weill A, Gabet A, Juillière Y, Cinaud A, Olié V.
379 Fourteen-Year Temporal Trends in Patients Hospitalized for Mitral Regurgitation: The
380 Increasing Burden of Mitral Valve Prolapse in Men. *J Clin Med Res* [Internet]. 2022;11.
381 Available from: <http://dx.doi.org/10.3390/jcm11123289>
- 382 25. Mitchell E, Walker R. Global ageing: successes, challenges and opportunities. *Br J Hosp*

- 383 *Med* . 2020;81:1–9.
- 384 26. Zhang J, Gajjala S, Agrawal P, Tison GH, Hallock LA, Beussink-Nelson L, Lassen MH,
385 Fan E, Aras MA, Jordan C, Fleischmann KE, Melisko M, Qasim A, Shah SJ, Bajcsy R, Deo
386 RC. Fully automated echocardiogram interpretation in clinical practice. *Circulation*.
387 2018;138:1623–1635.
- 388 27. Zoghbi WA, Adams D, Bonow RO, Enriquez-Sarano M, Foster E, Grayburn PA, Hahn RT,
389 Han Y, Hung J, Lang RM, Little SH, Shah DJ, Shernan S, Thavendiranathan P, Thomas JD,
390 Weissman NJ. Recommendations for noninvasive evaluation of native valvular
391 regurgitation. *J Am Soc Echocardiogr*. 2017;30:303–371.
- 392 28. Tran D, Wang H, Torresani L, Ray J, LeCun Y, Paluri M. A Closer Look at Spatiotemporal
393 Convolutions for Action Recognition [Internet]. arXiv [cs.CV]. 2017 [cited 2023 Oct
394 19];Available from: <http://arxiv.org/abs/1711.11248>
- 395 29. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: Visual
396 explanations from deep networks via gradient-based localization. *Int J Comput Vis*.
397 2020;128:336–359.
- 398 30. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot
399 when evaluating binary classifiers on imbalanced datasets. *PLoS One*. 2015;10:e0118432.
- 400 31. Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks [Internet]. arXiv
401 [cs.LG]. 2017 [cited 2023 Nov 4];Available from: <http://arxiv.org/abs/1703.01365>
- 402 32. Steffner K, Christensen M, Gill G, Bowdish M, Rhee J, Kumaresan A, He B, Zou J, Ouyang
403 D. Deep learning for transesophageal echocardiography view classification [Internet].
404 bioRxiv. 2023;Available from:
405 <https://www.medrxiv.org/content/10.1101/2023.06.11.23290759.abstract>
- 406 33. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of
407 echocardiograms using deep learning. *NPJ Digit Med* [Internet]. 2018;1. Available from:
408 <http://dx.doi.org/10.1038/s41746-017-0013-1>

- 409 34. Arnaout R, Curran L, Zhao Y, Levine JC, Chinn E, Moon-Grady AJ. An ensemble of neural
410 networks provides expert-level prenatal detection of complex congenital heart disease. *Nat*
411 *Med.* 2021;27:882–891.
- 412 35. Chiu I-M, Lin C-HR, Yau F-FF, Cheng F-J, Pan H-Y, Lin X-H, Cheng C-Y. Use of a Deep-
413 Learning Algorithm to Guide Novices in Performing Focused Assessment With Sonography
414 in Trauma. *JAMA Netw Open.* 2023;6:e235102.
- 415

Table 1 – View Classifier Demographics

417

	Train	Validation	Test
Videos, n (%)	67,179 (80.0)	8,344 (9.9)	8,454 (10.1)
Patients, n (%)	30,761 (80.0)	3,848 (10.0)	3,851 (10.0)
Studies, n (%)	46,882 (80.0)	5,790 (9.9)	5,942 (10.1)
Male, n (%)	38,367 (57.1)	4,907 (58.8)	4,854 (57.4)
Hypertension, n (%)	40,982 (61.0)	5,070 (60.8)	5,065 (59.9)
Coronary Artery Disease, n (%)	29,661 (44.2)	3,670 (44.0)	3,758 (44.5)
Atrial Fibrillation, n (%)	19,909 (29.6)	2,289 (27.4)	2,498 (29.5)
Ejection Fraction, mean (SD), %	56.56 (16.2)	57.11 (15.5)	55.85 (16.4)
Left Atrial Volume Index (LAVI), mean (SD), mL/m²	34.71 (15.4)	34.49 (15.4)	34.05 (15.1)
Race, n (%)			
White	47,110 (70.1)	5,927 (71.0)	5,709 (67.5)
Black	8,282 (12.3)	972 (11.6)	1,228 (14.5)
Asian	5,232 (7.8)	579 (6.9)	640 (7.6)
Other	6,555 (9.8)	866 (10.4)	877 (10.4)
MR Severity, n (%)			
No MR	24,840 (37.0)	3,107 (37.2)	3,291 (38.9)
Mild	25,235 (37.6)	3,143 (37.7)	3,104 (36.7)
Moderate	12,477 (18.6)	1,519 (18.2)	1,518 (18.0)
Severe	4,627 (6.9)	575 (6.9)	541 (6.4)

418 **Table 1** - Clinical and demographic characteristics represented in the training, validation, and
 419 internal test data sets for the 83,977 apical-4-chamber videos used to train, validate, and test the
 420 mitral doppler A4C view classifier. Values outside and inside parentheses represent number and
 421 percent, respectively, for categorical variables and mean and standard deviation for continuous
 422 variables.

423 **Table 2 – Mitral Regurgitation Cohort Demographics**

424

	Train	Validation	Test
Videos, n (%)	16,533 (80.2)	2,007 (9.7)	2,064 (10.0)
Patients, n (%)	11,623 (80.0)	1,450 (10.0)	1,450 (10.0)
Studies, n (%)	14,565 (80.3)	1,768 (9.8)	1,800 (9.9)
Male, n (%)	9,628 (58.2)	1,143 (57.0)	1,152 (55.8)
Hypertension, n (%)	10,087 (61.0)	1,179 (58.7)	1,296 (62.8)
Coronary Artery Disease, n (%)	7,482 (45.3)	888 (44.2)	963 (46.7)
Atrial Fibrillation, n (%)	5,414 (32.7)	599 (29.8)	642 (31.1)
Ejection Fraction, mean (SD), %	54.53 (17.6)	55.26 (17.5)	54.0 (17.9)
Left Atrial Volume Index (LAVI), mean (SD), mL/m²	36.85 (15.8)	36.35 (15.8)	37.44 (16.3)
Race, n (%)			
White	11,567 (70.0)	1,378 (68.7)	1,400 (67.8)
Black	1,983 (12.0)	277 (13.8)	283 (13.7)
Asian	1,428 (8.6)	147 (7.3)	192 (9.3)
Other	1,555 (9.4)	205 (10.2)	189 (9.2)
MR Severity, n (%)			
No MR	4,949 (29.9)	633 (31.5)	624 (30.2)
Mild	4,990 (30.2)	606 (30.2)	586 (28.4)
Moderate	4,952 (30.0)	595 (29.6)	627 (30.4)
Severe	1,642 (9.9)	173 (8.6)	227 (11.0)

425
 426 **Table 2** - Clinical and demographic characteristics of the 20,604 apical-4-chamber videos used
 427 to train, validate, and test the MR model. Values outside and inside parentheses represent
 428 number and percent, respectively, for categorical variables and mean and standard deviation for
 429 continuous variables.

430

431 **Table 3 - Model Performance**

<u>Site</u>	<u>Class</u>	<u>AUC</u>	<u>PPV</u>	<u>NPV</u>	<u>Recall</u>	<u>F1-Score</u>
CSMC	≥ Moderate MR	0.916 (0.899 - 0.932)	0.805 (0.766 - 0.842)	0.863 (0.835 - 0.890)	0.807 (0.769 - 0.843)	0.806 (0.776 - 0.834)
	Severe MR	0.934 (0.913 - 0.953)	0.626 (0.533 - 0.713)	0.954 (0.940 - 0.967)	0.626 (0.535 - 0.713)	0.626 (0.546 - 0.695)
Stanford	≥ Moderate MR	0.951 (0.924 - 0.973)	0.509 (0.427 - 0.591)	0.986 (0.974 - 0.995)	0.930 (0.867 - 0.983)	0.658 (0.581 - 0.727)
	Severe MR	0.969 (0.946 - 0.987)	0.674 (0.537 - 0.809)	0.977 (0.962 - 0.990)	0.729 (0.580 - 0.860)	0.701 (0.581 - 0.800)

432

433 **Table 3 - Model performance across institutions** - AUC, PPV, NPV, Recall and F1-score for
 434 MR on an internal test set from CSMC and an external validation set at Stanford. 95%
 435 confidence intervals were obtained by bootstrapping 10,000 samples. “≥ moderate” includes
 436 moderate and severe MR.

437

438

439

440

441

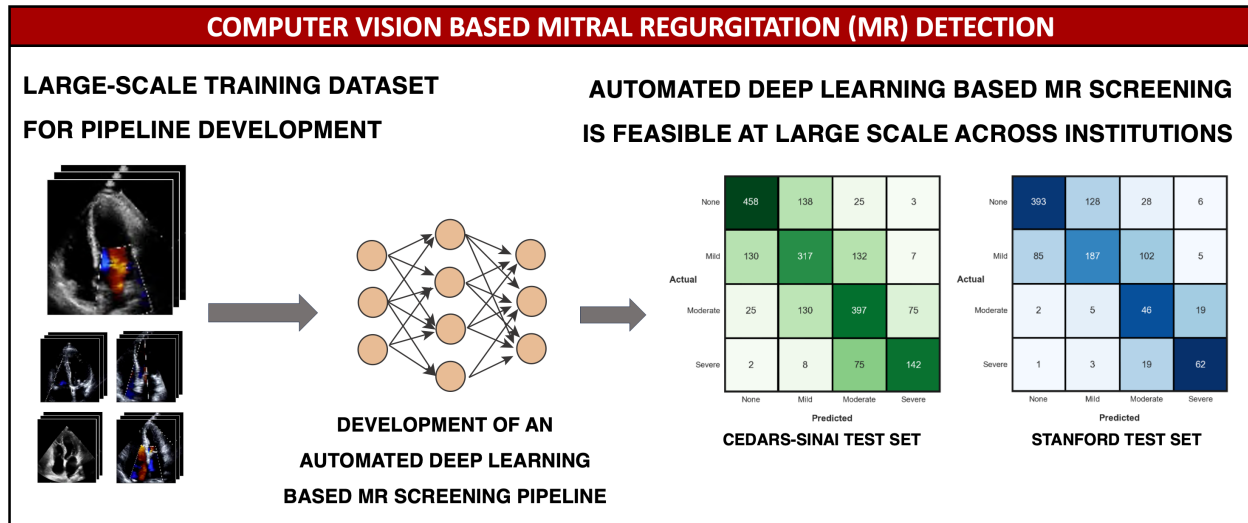
442

443

444

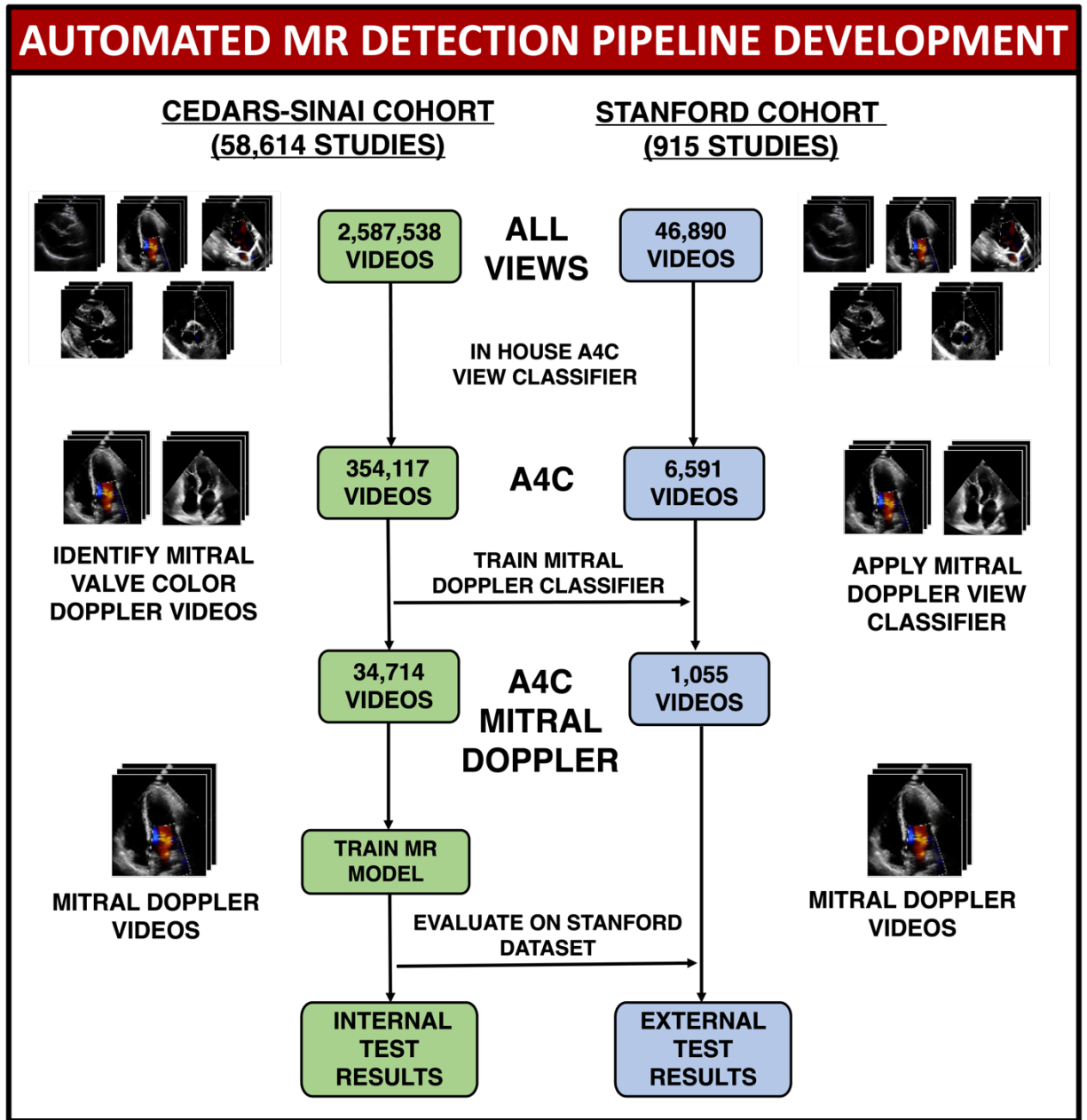
445

446



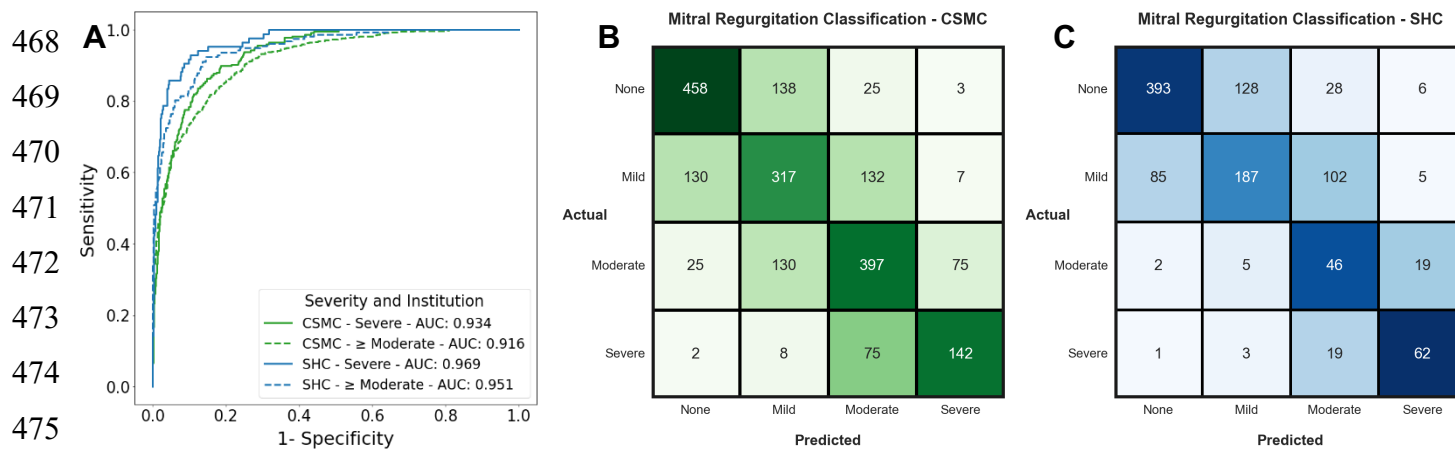
447
448
449
450
451
452
453
454
455
456
457
458
459

Figure 1: Computer Vision Based Mitral Regurgitation (MR) Detection: An automated deep learning pipeline was trained to detect and stratify mitral regurgitation severity using large-scale data consisting of apical-4-Chamber (A4C) echocardiogram videos with Color Doppler across the mitral valve (CSMC). The automated pipeline showed strong and consistent performance in test sets at CSMC and SHC. These results show that deep learning can accurately detect clinically significant MR using single-view TTE videos with Doppler information. Deep learning-based MR detection tools could serve as a part of point-of-care ultrasound screening as part of clinic visits or in resource limited settings where imaging may be obtained by individuals with minimal training.



460
461
462
463
464
465
466
467

Figure 2: CSMC and Stanford Dataset Isolation. 34,714 Color Doppler A4C videos were isolated from a larger set of videos from CSMC. A view classifier was trained and used to isolate A4C Mitral Doppler videos from 915 studies containing 1,055 suitable videos from Stanford Healthcare. The MR classification model was then benchmarked on an internal test set from CSMC and an external test set from Stanford Healthcare.



477 **Figure 3: Model Performance Across Severity and Institution – A.** Receiver operating
 478 characteristic (ROC) curves for detection of Severe or \geq Moderate MR at CSMC and Stanford.
 479 “ \geq Moderate” included moderate, moderate to severe, and severe MR. **3B and 3C:** MR
 480 Classification on test set videos from CSMC and Stanford, respectively. Confusion matrix
 481 colormap values were scaled based on the proportion of actual disease cases in each class that
 482 were predicted in each possible disease category. This was done to allow for relative comparison
 483 of model performance across disease classes (None, Mild, Moderate, and Severe) given class
 484 imbalance.
 485

486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522

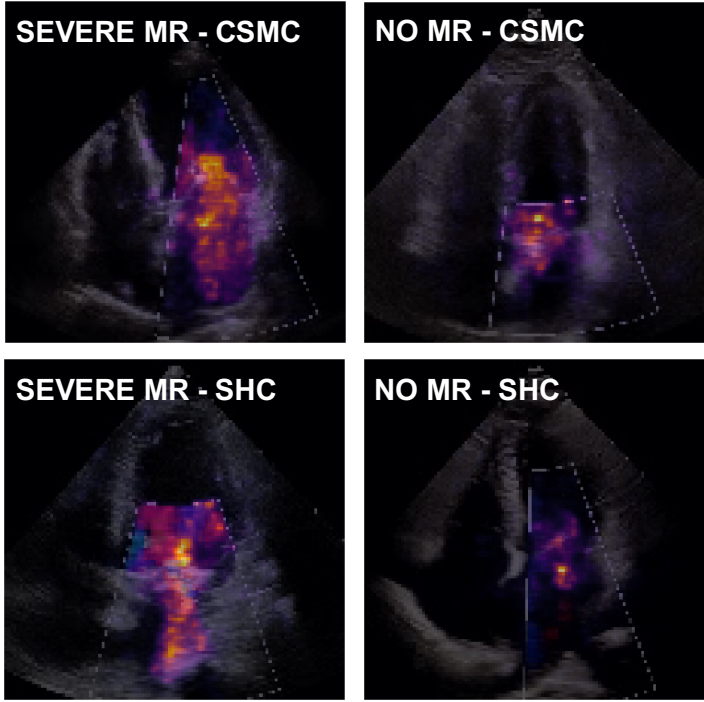


Figure 4: Saliency Map Visualization for MR classification models. Echocardiogram videos with severe MR from CSMC (top left) and Stanford (bottom left) are shown on the left, while videos with no MR from CSMC (top right) and Stanford (bottom right) are shown on the right. Saliency maps were computed using the Integrated Gradients method. A final 2-dimensional heatmap was generated by using the maximum value along the temporal axis for each pixel location in the video. Pixels brighter in color and closer to yellow were more salient to model predictions, while those darker in color were less important to the model’s final prediction. Severe MR was assessed by using the activation function for severe disease to generate a heatmap. When assessing controls, heatmaps were generated by stacking heatmaps for severe and moderate classes and taking the maximum between the two at each pixel location.