JACC: CARDIOVASCULAR IMAGING
VOL. 18, NO. 1, 2025

PUBLISHED BY ELSEVIER

ORIGINAL RESEARCH

An Automated Machine Learning–Based Quantitative Multiparametric Approach for Mitral Regurgitation Severity Grading

Anita Sadeghpour, MD,[a] Zhubo Jiang,[b] Yoran M. Hummel, PhD,[b] Matthew Frost, BE,[b] Carolyn S.P. Lam, MD,[c] Sanjiv J. Shah, MD,[d] Lars H. Lund, MD,[e] Gregg W. Stone, MD,[f] Madhav Swaminathan, MD,[g] Neil J. Weissman, MD,[a] Federico M. Asch, MD[a]

ABSTRACT

BACKGROUND Considering the high prevalence of mitral regurgitation (MR) and the highly subjective, variable MR severity reporting, an automated tool that could screen patients for clinically significant MR (≥ moderate) would streamline the diagnostic/therapeutic pathways and ultimately improve patient outcomes.

OBJECTIVES The authors aimed to develop and validate a fully automated machine learning (ML)–based echocardiography workflow for grading MR severity.

METHODS ML algorithms were trained on echocardiograms from 2 observational cohorts and validated in patients from 2 additional independent studies. Multiparametric echocardiography core laboratory MR assessment served as ground truth. The machine was trained to measure 16 MR-related parameters. Multiple ML models were developed to find the optimal parameters and preferred ML model for MR severity grading.

RESULTS The preferred ML model used 9 parameters. Image analysis was feasible in 99.3% of cases and took 80 $\pm$ 5 seconds per case. The accuracy for grading MR severity (none to severe) was 0.80, and for significant (moderate or severe) vs nonsignificant MR was 0.97 with a sensitivity of 0.96 and specificity of 0.98. The model performed similarly in cases of eccentric and central MR. Patients graded as having severe MR had higher 1-year mortality (adjusted HR: 5.20 [95% CI: 1.24-21.9]; $P = 0.025$ compared with mild).

CONCLUSIONS An automated multiparametric ML model for grading MR severity is feasible, fast, highly accurate, and predicts 1-year mortality. Its implementation in clinical practice could improve patient care by facilitating referral to specialized clinics and access to evidence-based therapies while improving quality and efficiency in the echocardiography laboratory. (JACC Cardiovasc Imaging. 2025;18:1-12) © 2025 by the American College of Cardiology Foundation.

Mitral regurgitation (MR) is the most prevalent valvular heart disease in the United States and the second-most common in Europe.[1-3] As newer technologies become available for treating patients with significant MR,[4-6] the need for proper diagnosis becomes increasingly relevant.

Echocardiography is the primary imaging modality recommended for determining the mechanism and

From the [a]MedStar Health Research Institute and Georgetown University, Washington, District of Columbia, USA; [b]Us2.ai, Singapore, Singapore; [c]National Heart Centre Singapore, Duke-National University of Singapore, Singapore; [d]Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA; [e]Karolinska University Hospital, Stockholm, Sweden; [f]The Zena and Michael A. Wiener Cardiovascular Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA; and the [g]Department of Anesthesiology, Duke University School of Medicine, Durham, North Carolina, USA.

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received February 12, 2024; revised manuscript received May 16, 2024, accepted June 20, 2024.

ISSN 1936-878X/$36.00
https://doi.org/10.1016/j.jcmg.2024.06.011

JACC: CARDIOVASCULAR IMAGING, VOL. 18, NO. 1, 2025
JANUARY 2025:1-12

ABBREVIATIONS AND ACRONYMS

A2C = apical 2-chamber

A4C = apical 4-chamber

CWDD = continuous wave Doppler jet density

LA = left atrial

LV = left ventricular

LVEDV = left ventricle end-diastolic volume

LVESV = left ventricle end-systolic volume

LVOT = left ventricular outflow tract

MR = mitral regurgitation

ML = machine learning

PASP = pulmonary artery systolic pressure

PLAX = parasternal long-axis

ROI = region of interest

RAR = color Doppler regurgitant jet area ratio

TTE = transthoracic echocardiograms

VC = vena contracta width

severity of MR.[2] Given the complex anatomy and function of the mitral valve (MV) apparatus, grading MR severity is challenging, requiring a comprehensive examination with multiple parametric assessments.[7] Despite notable advances in the understanding of the MV anatomy and function, the integrated multiparametric approach suggested by current guidelines is time-consuming and subject to significant interobserver variability,[8] mainly related to subjective evaluation of color Doppler and other proposed measurements, which affects proper treatment selection and timing for interventions. Even newer quantitative measures of MR have proven to be cumbersome, have slow adoption in the community, and result in high variability, thus leaving the clinician with doubt as to the true severity of MR.

Recent studies have demonstrated the promising role of artificial intelligence (AI) and machine learning (ML) in cardiac imaging.[9,10] In cardiac ultrasound, new AI technologies are being developed to help in image acquisition,[11] as well as automated cardiac chamber size and function assessment.[12-14] Automated analysis makes the reporting of an echocardiogram significantly more efficient while improving the reproducibility of measurements such as left ventricular (LV) volumes and ejection fraction (EF).[12,15-18] AI's role can expand to diagnosing other cardiovascular conditions, such as valvular heart disease.

Categorizing MR severity is essential because significant MR is associated with high morbidity and mortality.[19,20] An automated tool that could screen patients for significant MR would improve timely disease detection, make this diagnosis more reliable, and streamline the diagnostic and therapeutic pathways, ultimately improving patient outcomes.

We, therefore, developed a fully automated ML-based workflow to quantify parameters of MR severity grading. Here, we describe the development and validation of algorithms that aim to grade MR severity accurately by using transthoracic echocardiograms (TTE) from patients with heart failure and various degrees of MR severity included in multicenter clinical studies.

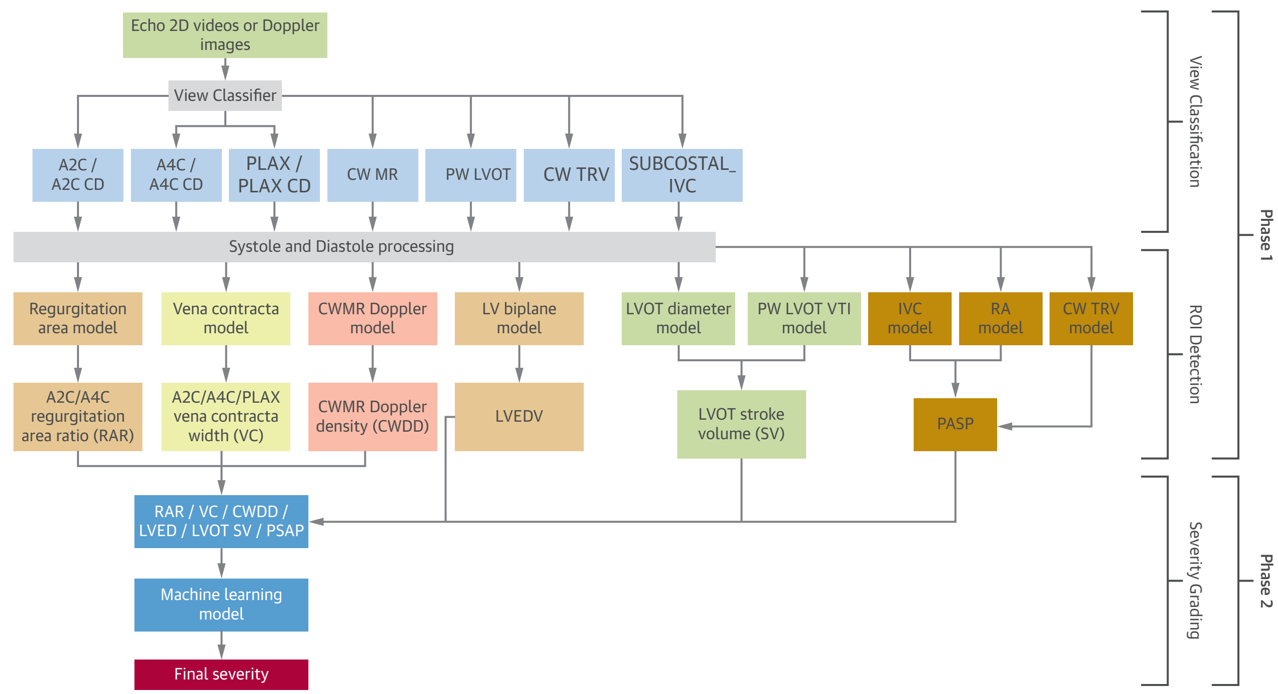
METHODS

To evaluate MR severity, we developed automated ML-based algorithms to measure 16 American Society of Echocardiography (ASE)-recommended echocardiographic MR-related parameters within 3 categories (Supplemental Table 1)[7]: MR-specific, chamber size, and hemodynamics parameters. The MR-specific quantitative parameters were as follows: MR jet vena contracta width (VC) measured in the parasternal long-axis (PLAX), apical 2-chamber (A2C), and apical 4-chamber (A4C) views, color Doppler regurgitant jet area ratio (RAR) measured in the A4C and A2C views, and the continuous wave Doppler jet density (CWDD). Image analysis algorithms were developed (study phase 1, measurements development stage), and later MR severity grading models were created (study phase 2, measurements validation and severity models development stage) in independent TTE data sets.

STUDY COHORTS. For phase 1, randomly selected TTEs of patients with any degree of MR enrolled in the previously described MacKay and ATTRaCT (Asian Network for Translational Research and Cardiovascular Trials) cohorts were used to develop ML algorithms for DICOM (digital imaging and communications in medicine) image analysis.[12] In brief, the ATTRaCT data set contains data from 11 countries involved in the Asian Sudden Cardiac Death in Heart Failure registry (heart failure [HF] with preserved or reduced left ventricular ejection fraction [LVEF]) and Mackay contains real-world patients from a large Hospital in Taipei, Taiwan. The phase 1 development data set was split into independent, randomly selected training and testing sets (90:10 ratio). The number of images or videos used to develop each of the algorithms is included in Supplemental Table 2.

For phase 2, a total of 438 patients randomly selected from 2 multicenter clinical studies (PROMIS-HFpEF [PRevalence Of MIcrovascular dySfunction in Heart Failure with Preserved Ejection Fraction] and COAPT [Cardiovascular Outcomes Assessment of the MitraClip Percutaneous Therapy for Heart Failure Patients With Functional Mitral Regurgitation]) were included. PROMIS-HFpEF included patients with HFpEF regardless of the presence, degree, or etiology of MR, whereas COAPT included patients with HF, LVEF 20%-50%, and moderate-to-severe or severe secondary MR.[5,21] The COAPT group included a subgroup of trial screen failures with moderate MR. As a result, the validation cohort included patients with various degrees of MR and LVEF. The phase 2 cohort was split into 2 independent subcohorts: the measurements external validation group (70%; n = 305), which was also used to develop the multiparametric models; and the independent testing group (30%; n = 133), which was used to test the performance of

FIGURE 1 Study Workflow for Algorithms Development (Phase 1) and Severity Grading (Phase 2) Validation Phases



The multiparametric MR severity is calculated using a series of ML-derived models that include the 3 MR-related echocardiography categories: 1) MR-specific parameters: RAR, VC, and CWDD; 2) left-sided heart chamber size parameters; and 3) hemodynamic parameters such as LVOT SV and PASP. Each of these parameters is calculated considering all available echocardiographic views, all videos and cardiac cycles within each view, and measurements from all frames of each cardiac cycle. 2D = 2-dimensional; A2C = apical 2-chamber; A4C = apical 4-chamber; CD = color Doppler; CW = continuous wave; CWDD = continuous wave Doppler jet density; IVC = inferior vena cava; LV = left ventricle; LVOT = left ventricular outflow tract; LVEDV = left ventricle end diastolic volume; ML = machine learning; MR = mitral regurgitation; PASP = pulmonary artery systolic pressure; PLAX = parasternal long-axis; PW = pulse wave Doppler; RA = right atrium; RAR = regurgitation area ratio; RV = regurgitant volume; SV = stroke volume; TRV = tricuspid regurgitation velocity; VC = vena contracta; VTI = velocity time integral.

the single parameters and multiparametric models in grading MR severity.

Clinical studies were approved by the institutional review board/ethics committees and patient consent was obtained as appropriate.

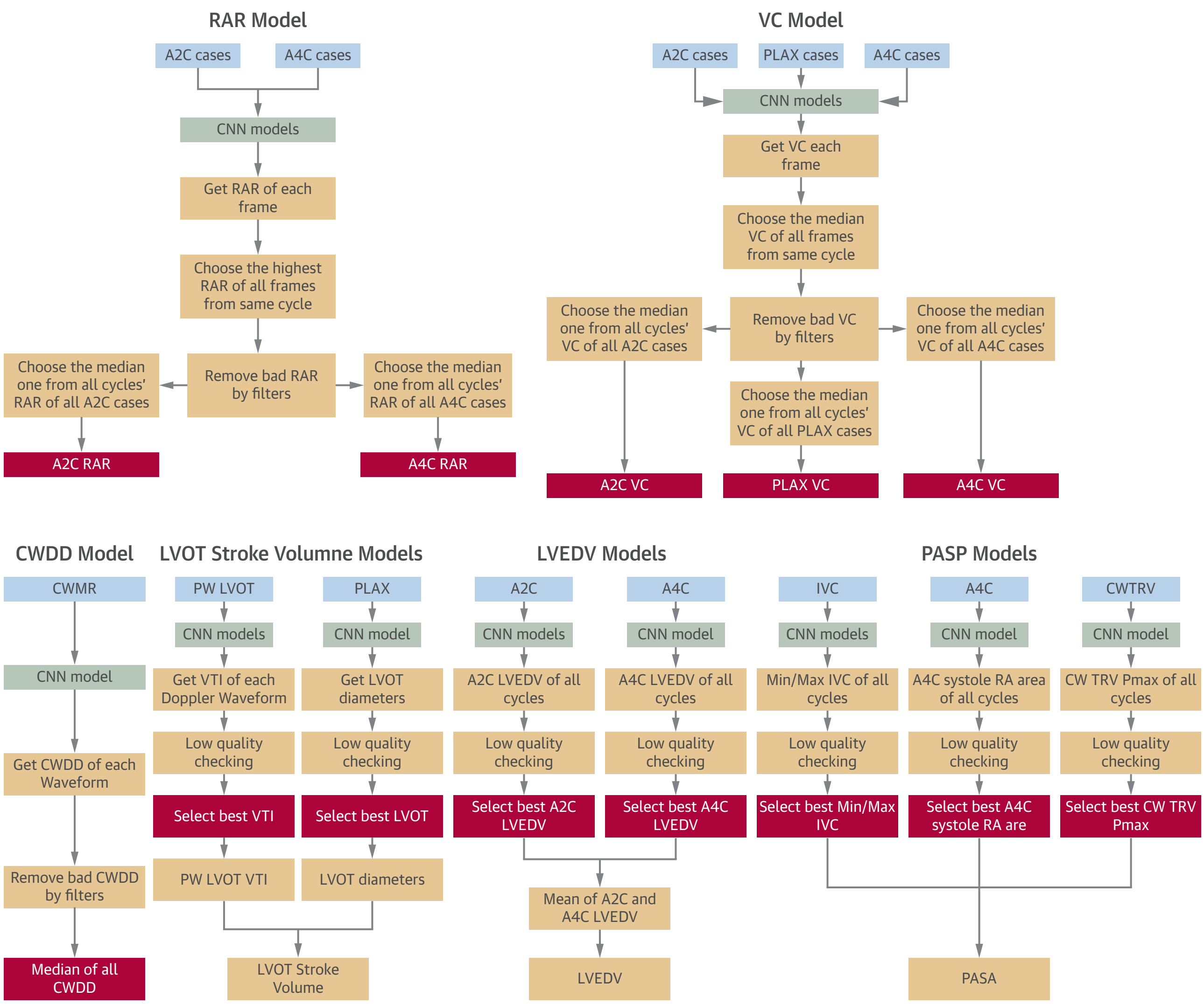
PHASE 1: MEASUREMENT ALGORITHMS DEVELOPMENT, IMAGE ANALYSIS TRAINING, AND TESTING. Echocardiographic views were classified (Figure 1) using the previously described automated view classifier.[12] For MR-specific parameters, 3 certified echocardiography experts employed by Us2.ai generated the region of interest (ROI) (2 experts generated ROI as training data, 1 reviewed each of them). Training and test data from ATTrACT were generated to train MR jet area, VC, and MR CW Doppler waveform models, selecting only high-quality images. Automatic annotations were created through a convolutional neural network to generate the ROI. Afterward, these models were used for MR detection in the Mackay cohort. Experts corrected any inaccurate detections from AI-generated MR detection, which were used to further improve the trained models. The experts subsequently performed manual annotations, which were used to train the machine in tracing each parameter's measurement.

The VC was measured in 3 views (PLAX, A4C, and A2C). All available cardiac cycles on the views of interest were used. The machine was trained to perform tracings and measurements for every systolic frame in the cardiac cycles in all video clips available for each view and to calculate the median for each clip (Figure 2). Because multiple clips/cardiac cycles were available for each view, the median of all of them was used as the final value for a given view. The VC parameter was considered separately for each of the 3 views and as a single final VC value (the maximum of the measurements obtained from each of the 3 views).

For RAR, the MR jet and left atrial areas were measured in all the available A2C and A4C views with

JACC: CARDIOVASCULAR IMAGING, VOL. 18, NO. 1, 2025
JANUARY 2025:1-12

FIGURE 2 Algorithm Development Process for Each Parameter

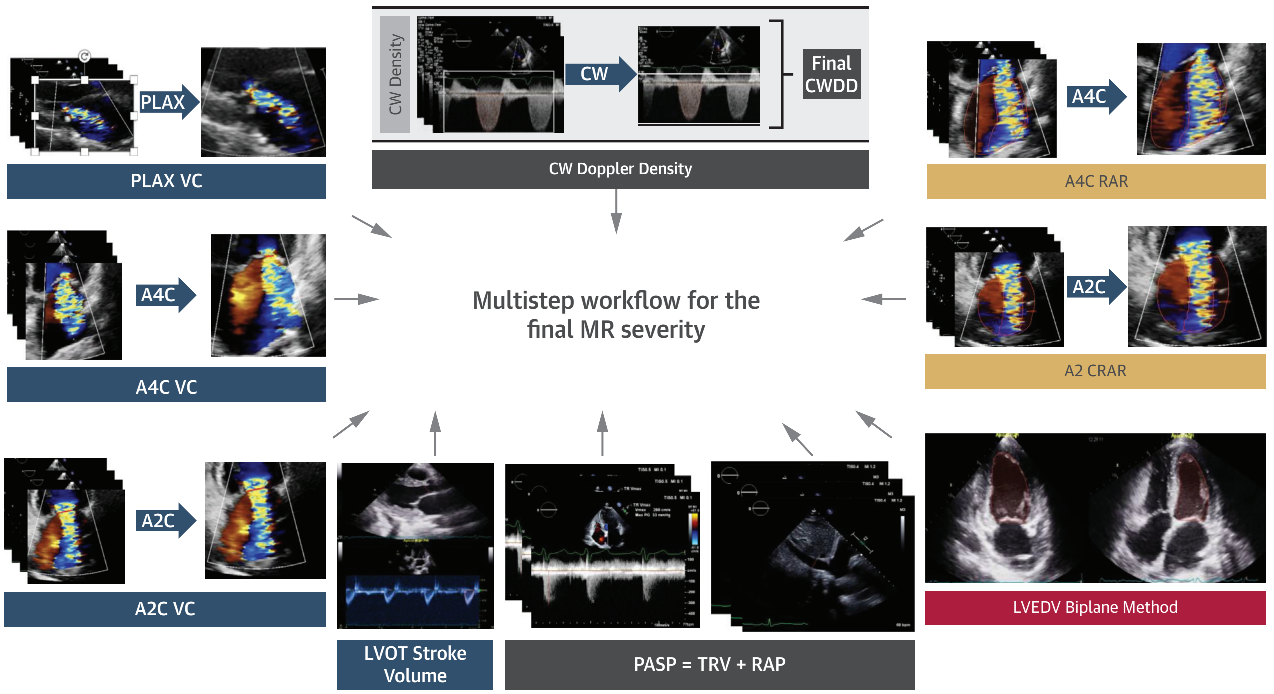


On each video, multiple cardiac cycles may be available. For the MR-specific parameters such as VC, RAR, and CWDD, all frames in a cycle are measured and the largest (RAR) or median (VC) one is selected, then the median for all cycles in each view is selected (final parameter for each view). Through all PW LVOT VTI images, the machine chose the highest PW maximum velocity. The best LVOT diameter was defined as the one with the highest annotation probability. For the LVEDV and LVESV, the echo 2D video with the highest view classification probability was selected, and the final LVEDV/LVESV was calculated from the cycle with the median of all cycles in the selected video. LVOT SV was calculated using the formula of 3.14 (LVOT diameter/2)^2 $\times$ LVOT VTI. PASP was calculated through TR peak velocity + right atrium pressure assessed using IVC size and collapse. If IVC was not visualized then RA area was used as RAa >18 cm^2, RAP = 10 mm Hg, and RA area <18 cm^2, and RAP 5 mm Hg. CNN = Convolutional Neural Network; other abbreviations as in Figure 1.

a color Doppler signal. A similar multiclip/cycle/frame analysis process was used for each view described already in this article (Figure 2). The RAR parameter was considered separately for each of the 2 views and as a single final RAR value (average of the 2 views).

The density of the MR CW signal, traditionally a subjective qualitative parameter, was measured as CWDD, a novel quantitative parameter. CWDD was analyzed by assigning a numerical value in the gray-scale to each pixel within the CW waveform. Because the greyscale can be altered during image acquisition, CWDD was obtained by normalizing to the brightness of the entire Doppler image (Supplemental Figure 1). CWDD values ranged from 0-1 (0 reflecting black,

FIGURE 3 Multiple Parameters Used for the Final MR Severity Grading Model



Multiple images were used for measuring the 9 selected parameters, including multiple videos of A4C views (for VC and RAR), A2C (for VC and RAR), PLAX (for VC), images with CW of the MR jet (for CWDD), as well as the videos to obtain the non-MR specific measurements (LVEDV, LVOT SV, and PASP). On each video multiple cardiac cycles may be available. Every systolic frame for each cycle is measured and the largest (RAR) or median (VC) one is selected. The median of all those selected for each view is the final measurement for the parameter in a given view. Once this process is completed for each view of interest for a given parameter (PLAX, A4C, and A2C for VC; A4C and A2C for RAR; all MR CW waveforms for CWDD; biplane LVEDV, LVOT SV, and PASP), these single-parameter values are combined in the preferred multiparametric model (model M7). RAP = right atrial pressure; other abbreviations as in Figure 1.

1 white, and shades of gray in between). All available cardiac cycles were used, and the median was selected as the final CWDD.

For chamber size and hemodynamic parameter analysis, a segmentation model was used to generate waveform masks of left ventricular outflow tract (LVOT) pulse wave Doppler signals and 2 heatmap regression models were developed to assess the inferior vena cava and LVOT diameters independently. CW tricuspid regurgitation peak velocity, and left atrial volume, MV peak E velocity, MV peak A velocity, E/A ratio, left ventricle end-diastolic volume (LVEDV), left ventricle end-systolic volume (LVESV), and right atrial area were measured using the models previously described[12,13] (Figure 3). All parameters were made following the ASE standards.[7]

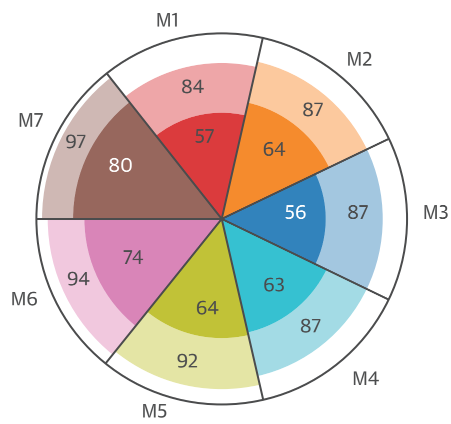
PHASE 2: MEASUREMENTS EXTERNAL VALIDATION AND MR SEVERITY GRADE MODELS DEVELOPMENT. Once the machine was trained for views selection and measuring all 16 parameters (phase 1), the measurements were used to grade MR severity by comparing to the ground truth in the MR severity validation phase (Figure 1). The echocardiograms were reviewed by experts in the MedStar Health core laboratory (ground truth) and graded as none/trace, mild, moderate, or severe (including moderate-to-severe) through a detailed qualitative and quantitative analysis following ASE Guidelines.[7] Finally, none/trace and mild were grouped as nonsignificant MR, and moderate and severe as significant MR. All core laboratory analysis was performed before the automated algorithms were developed and blinded to any clinical information.

The ML model performance was tested for each of the single MR-specific parameters (RAR, VC, CWDD models 1-3, respectively, in Supplemental Table 3) and various multiparametric models (models 4-7).

MULTIPARAMETRIC MODELS. Various models of multiparametric analysis were designed and tested using combinations of the 16 candidate parameters as model input and MR grade severity as output. The ML

FIGURE 4 Accuracy of All Models Developed for MR Severity Grading

M1: Single parameter ML performance for RAR
M2: Single parameter ML performance for VC
M3: Single parameter ML performance for CWDD
M4: Output is the median severity grade of the ML 1, 2 and 3
M5: Out put is based on the six MR specific parameters (PLAX VC, A2C VC, and A4C VC, A2C RAR, A4C RAR, CWDD) using ML decision tree model
M6: Output is based on the six MR specific parameters (PLAX VC, A2C VC, and A4C VC, A2C RAR, A4C RAR, CWDD) using the catboost model
M7: Output is based on the nine MR related parameters (PLAX VC, A2C VC, and A4C VC, A2C RAR, A4C RAR, CWDD, LVEDV, LVOT stroke volume, and PASP using the catboost model



Dark color mean all severity grades classification and light color means significant vs nonsignificant severity classification. Abbreviations as in Figure 1.

grading and measurement models were created with Catboost 1.2.2, TensorFlow 2.10.0, and Python 3.9. Specific descriptions of the 7 models are presented in Supplemental Table 3. A thorough analysis of the best parameters within the 3 specified groups was performed to find the optimal parameters from each group to enhance the overall model effectiveness.

CLINICAL OUTCOMES. One-year clinical outcomes (HF hospitalization and all-cause mortality) were collected from the PROMIS and COAPT trials to test the ML-preferred model's predictive value.

STATISTICAL ANALYSIS. The multiparametric MR grading assessment by the core laboratory (none/trace, mild, moderate, and severe) was considered the ground truth. The performance (sensitivity, specificity, and accuracy) of each single and multiparametric ML model was calculated. Accuracy, the overall probability that a patient is correctly classified, was calculated as:

$$\text{Accuracy} = \text{Sensitivity} \times \text{Prevalence} + \text{Specificity} \times (1 - \text{Prevalence})$$

Finally, the association of MR severity grade and 1-year clinical outcomes was tested using Cox proportional hazards models, adjusting to age and LVEF. Statistical analysis was done with SAS v9.4.

RESULTS

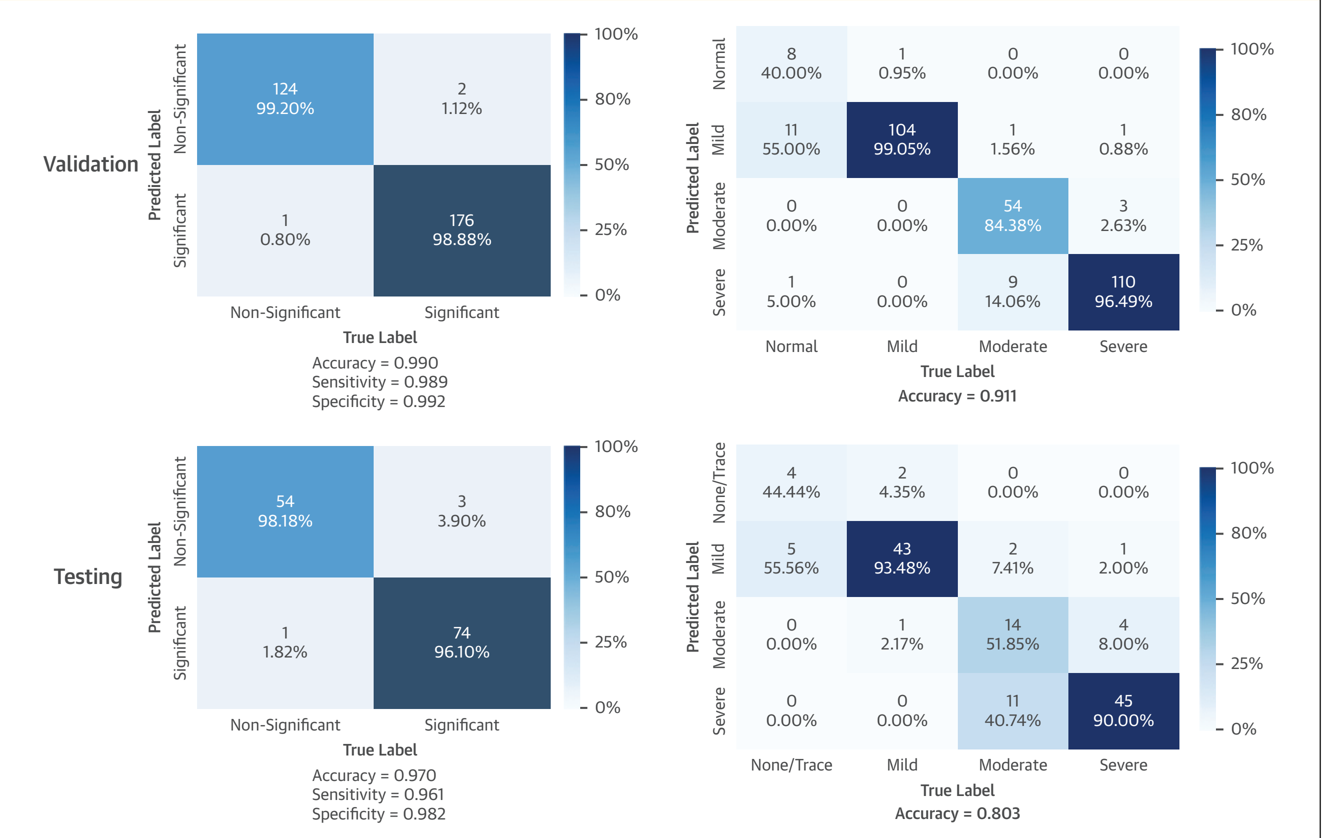
A total of 438 patients composed the population used for MR severity models development (phase 2). Basic demographic and echocardiographic characteristics are included in Supplemental Tables 4 and 5. The distribution of LVEDV, LVESV, LVEF, LA volume, right ventricular fractional area change, pulmonary artery systolic pressure (PASP), and MR severity (per core laboratory analysis) was broad, reflecting the diversity of the overall study population. There were no significant differences between patients included in the validation and testing cohorts.

Image analysis was feasible in 99.3% of the echocardiograms. Seven models were developed and tested; 3 used a single MR-specific parameter, and 4 were multiparametric.

The accuracy of MR severity grading for each MR-specific single parameter ranged from 0.56-0.64, and for significant vs nonsignificant MR ranged from 0.84-0.87. Specificities were slightly higher than sensitivities (Supplemental Table 6).

MR SEVERITY GRADING USING MULTIPARAMETRIC ANALYSIS. The use of multiple parameters in all tested ML models improved the MR severity grading as well as the identification of those with significant MR when compared with the single-parameter models. The performance of each of the ML models is described in Figure 4 and Supplemental Table 6. After testing multiple combinations through decision tree and catboost models, the best MR severity grading results were obtained with the ML catboost model (M7) using 9 independent selected parameters (Figure 3). Automated analysis was possible in 95.9% of cases for A2C/A4C RAR, 90.8% for A4C VC, 86.9% for A2C VC, 72.2% for PLAX VC, 87.8% for CWDD, 90.6% for PASP, 93.8% for LVEDV, and 89% for LVOT stroke volume (SV). The overall MR grade by automated analysis in model M7 was obtained by having

FIGURE 5 Performance of the Preferred ML Model for Grading All MR Severity and Nonsignificant (None, Trace, Mild) vs Significant (Moderate, Severe) MR in Validation and Testing Cohorts



Percentage values in each box refer to percentage of each category as determined by the core laboratory (true label, columns).

all 9 parameters in 27.1% of cases, at least 7 in 70.6%, and at least 5 in 85.1%. The mean time for full automated analysis (from image upload to full severity reporting) was 80 ± 5 seconds. Reproducibility was tested by running the models thrice in 40 randomly selected cases, with perfect agreement (0 variability).

Accuracy on the MR severity testing cohort for the multiparametric models ranged from 0.63-0.80 on grading severity and 0.87-0.97 on detecting significant vs nonsignificant MR.

Results on the performance of the preferred catboost model M7 were similar in the validation and testing cohorts (Figure 5). This model had an accuracy for grading MR severity of 0.80, whereas it had a specificity of 0.98, sensitivity of 0.96, and accuracy of 0.97 for detection of significant MR. The model performed with high accuracy in cases with eccentric and central MR jets (0.89 and 0.81, respectively) (Supplemental Figure 2).

ASSOCIATION OF MR SEVERITY AND CLINICAL OUTCOMES. Outcomes data were available for 351 patients. The 1-year all-cause mortality rate was 8.8% (n = 31), HF hospitalization was 13.7% (n = 48), and the rate of the combined endpoint of all-cause death and HF hospitalization was 18.5% (n = 65).

Based on the MR grading by the preferred model, death from any cause at 1 year occurred in 2.78% of patients with none or mild, 0% of those with moderate, and 15.8% of patients with severe MR. There was a significantly increased risk of 1-year death in those with severe compared with those with mild MR (HR: 6.44 [95% CI: 2.5-16.8]; $P < 0.001$), even after adjusting for age and LVEF (HR: 5.2 [95% CI: 1.24-21.9]; $P = 0.025$ when adjusted for LVEF and age by Cox proportional model). The combined endpoint of death and HF hospitalization occurred in 5.6%, 15.4%, and 33.5% of those with mild, moderate, and severe MR, respectively.

Based on the MR grading by the expert readers (ground truth) the increased risk of 1-year death for those graded severe MR was significant based on univariate analysis (HR: 5.91 [95% CI: 2.26-15.45]; $P = 0.0003$), but not statistically significant once adjusted for age and LVEF (HR: 3.66 [95% CI: 0.82-16.25]; $P = 0.088$).

DISCUSSION

In the current study, we developed and validated multiple ML-based algorithms for a novel, fully automated analysis of echocardiographic images for grading MR severity through single and multiparametric models. We leveraged the strengths of AI to generate and analyze large data sets in a very fast time (roughly 80 s) by automatically performing measurements from multiple cardiac cycles and multiple echocardiographic views. Our main findings are as follows: 1) the development of CWDD, a novel ML-based quantitative CW Doppler metric to grade MR severity, and automated models for measuring the MR guidelines-recommended parameters, VC, MR jet area ratio, LVEDV, PASP, and LVOT SV; 2) the development of several novel ML models for multiparametric evaluation of MR severity; 3) proof of high feasibility, efficiency, and accuracy of the automated ML workflow to classify the TTE views, detect MR jets, and perform quantification and grading of the MR severity by using multiple quantitative parameters and multiparametric models (**Central Illustration**); and 4) the significant association between the ML-predicted probability of MR severity and clinical outcomes.

During the last decade, the use of ML and AI in cardiovascular medicine has grown exponentially. In cardiac imaging and echocardiography, in particular, the application of algorithms to analyze big data and the need to improve the workflow, efficiency, and accuracy of the echocardiographic readings have been of particular interest. Several large studies have shown the feasibility and accuracy of automated analysis for assessing LV size and function, which is currently being applied in clinical trials[16] and daily clinical practice.[12,15,22]

MR evaluation is complex and often an area of uncertainty in clinical practice. For more than 2 decades, quantitative parameters have been introduced and promoted to improve the accuracy and reliability of MR evaluation. However, these quantitative measures are often difficult to apply in clinical practice, have high variability, and can be discordant between different quantitative measures. Despite the introduction of new MR measurements, these limitations have resulted in a very high degree of variability and decreased confidence in MR evaluations in the clinical setting.[23] These limitations have been recognized in recent guidelines that called for "more automation in quantitation [of valvular regurgitation] to reduce variability."[7]

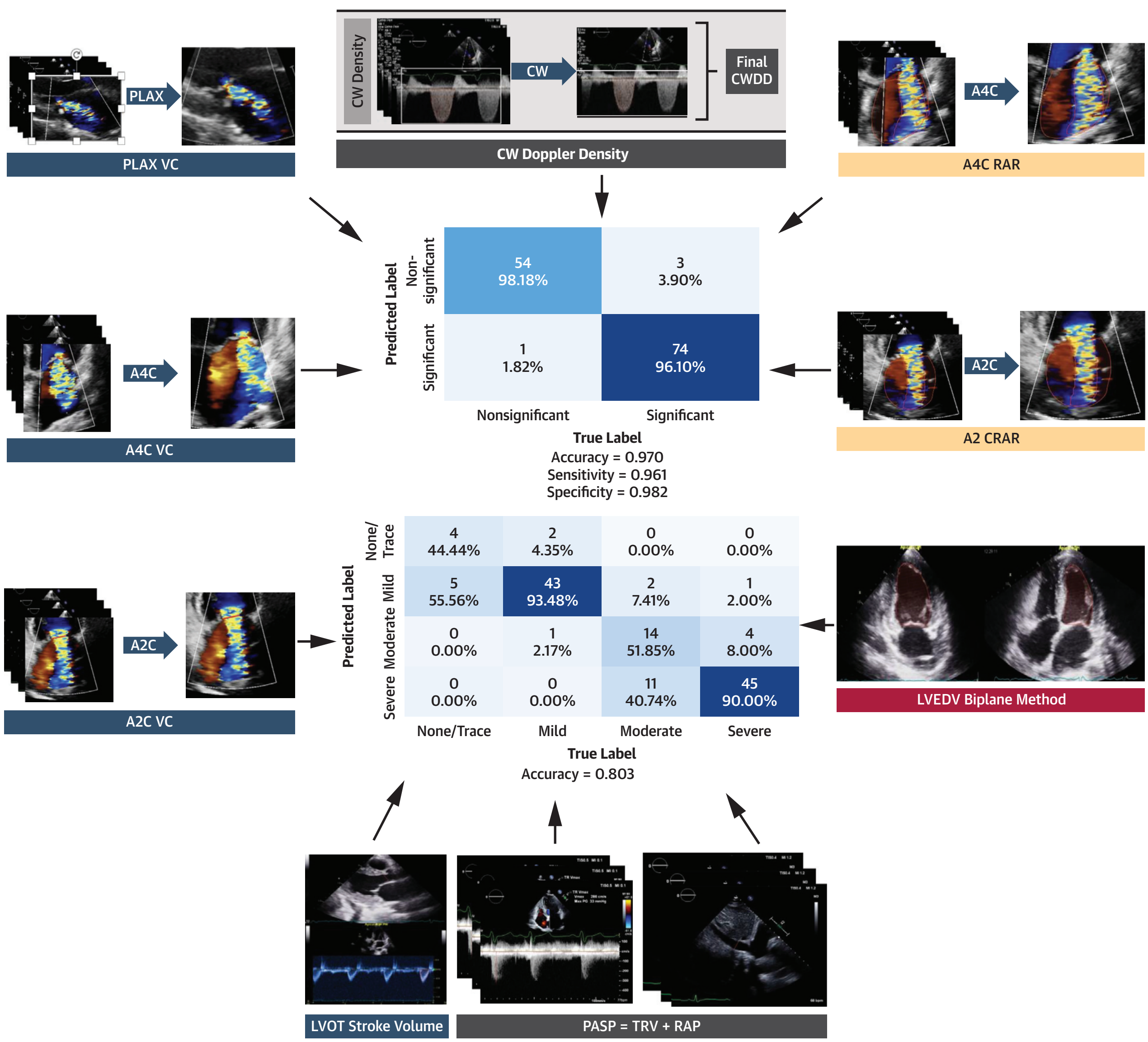
Despite the high interest from the medical community and AI's great potential for timely diagnosis and management of patients with valvular disease, few studies have explored its value in grading MR severity. Most recent advances in the application of AI in echocardiography assessment of valvular disease have been on their anatomic evaluation.[22,24,25]

In 2 AI-based MR severity assessment studies, automated models showed a high degree of accuracy for MR severity grading, but they used overly simplistic models involving only color Doppler imaging.[26,27] Our study differs from them in that we addressed the complexity of MR by comprehensively using color and spectral Doppler methods as well as LV size and hemodynamic parameters obtained in multiple cardiac cycles and from multiple views, and by testing against a core laboratory with expertise in detailed MR severity grading for clinical trials. In addition, our algorithms proved to perform best when using multiparametric ML models, a feature unique to our study and compliant with the ASE guidelines recommendations.[7] A recent study reports high feasibility in view classification but, different from our report, focused on the presence or absence of MR only in the PLAX view as the screening method for rheumatic mitral disease in children.[28] Our automated ML workflow differed from their study by not only detecting but also grading the MR severity in multiple views and including patients with various degrees and etiologies of MR.

In a field that is rapidly growing, the most novel aspects of our work are the quantitation of CWDD and the creation of the ML models that resulted in the development of a comprehensive MR evaluation model including 9 automated measurements that has high sensitivity, specificity, and accuracy.

Even at the current time of expanding therapeutic options, MR remains underdiagnosed and undertreated even in communities and centers with readily available diagnostic and treatment options.[29] Delays in diagnosis and proper management of significant MR can lead to poor outcomes, including HF, costly and recurrent hospital admission, and death. The current practice for MR evaluation requires a comprehensive, integrated approach including clinical information, valve morphology, etiology, and mechanism of MR in addition to the combined qualitative and quantitative MR

CENTRAL ILLUSTRATION Multiparametric Workflow for the MR Severity Grading Using Preferred ML Model



Sadeghpour A, et al. JACC Cardiovasc Imaging. 2025;18(1):1-12.

The performance of the selected 9 independent MR-related parameters as the best echocardiographic parameters for the MR severity grading and nonsignificant (none, trace, mild) vs significant (moderate, severe) MR using the preferred ML model is shown. Percentage values in each box refer to percentage of each category as determined by the core laboratory (true label, columns). A2C = apical 2-chamber; A4C = apical 4-chamber; CW = continuous wave; CWDD = continuous wave Doppler jet density; LVOT = left ventricular outflow tract; LVEDV = left ventricle end diastolic volume; MR = mitral regurgitation; PASP = pulmonary artery systolic pressure; PLAX = parasternal long-axis; RAR = regurgitation area ratio; RAP = right atrium pressure; SV = stroke volume; TRV = tricuspid regurgitation velocity; VC = vena contracta.

severity assessment. In this context, adopting AI technologies in echocardiography could improve diagnosis at a large scale and, therefore, improve patient screening to provide better care. Specifically for MR, the concept of a fast, accurate, and reproducible technology for diagnosis, such as the models presented here, could have significant implications in the workflow of echocardiography laboratories by improving reproducibility and quality performance. Our proposed multiparametric model had an excellent performance for grading severity, and, most importantly, it had almost perfect accuracy for detection of significant MR, a distinction that could benefit patients by the rapid identification of those that would need further evaluation, should be referred to a specialized valve or heart failure clinic, be considered for specific therapies, or, otherwise, identify those with no or mild MR who have benign clinical course.[30] We believe our proposed approach to MR grading could bring "echo interpretation expertise" to the community to facilitate proper referral to specialized clinics and centers, improving patient care.

STUDY LIMITATIONS. Our study was conducted retrospectively and involved echocardiograms from a specific group of patients (those with HF). By including patients from 2 different clinical studies, however, the study population was diverse and better characterized, particularly regarding the echocardiographic evaluation by a recognized core laboratory that served as the ground truth for MR severity grading. Nevertheless, patients with HF across the EF spectrum often have secondary MR, which may not represent the full spectrum of patients with MR. Specifically, those with primary MR from leaflet prolapse/flail mostly have eccentric MR jets, which can result, eg, in poor Doppler alignment of CW Doppler, lower MR jet density on CW imaging, and, therefore, may be underestimated by the CWDD algorithm. Our cohort, however, included patients with eccentric jets due to asymmetric leaflet tethering (40% of our cohort had ischemic MR), and the model performed similarly in those with central and eccentric jets. Finally, our study was not designed to analyze clinical outcomes related to automated AI-based diagnosis of MR severity. Therefore, the positive relationship between ML-predicted MR severity and clinical outcomes in our study should be considered hypothesis-generating. The design of proper clinical trials in AI and echocardiography with outcomes reporting is desirable[22] and should be the focus in the future. However, before clinical trials are designed to determine the clinical impact of AI-based echocardiography, automated AI-based methods of measuring echocardiographic parameters and detecting cardiovascular diseases must be developed, as we have done in this study. Similarly, the application of these multiparametric models in large cohorts of patients from health systems (real world) or research registries will be needed to prove the ultimate value of our approach in the general population and everyday practice.

CONCLUSIONS

A novel, fully automated ML multiparametric model for grading MR severity using guideline-recommended parameters is feasible, highly accurate, and predicts 1-year mortality. Its implementation in clinical practice could improve echocardiography laboratory workflow and quality, ultimately improving patient care by facilitating appropriate referrals to specialized clinics and centers and by timely offering specific therapies.

ACKNOWLEDGMENTS The authors thank Biniyam G. Demissei and Stephen Fernandez for statistical support and Camilla Hage, Abbott (Saurabh Datta and Barathi Sethuraman), and AstraZeneca for sharing data from the COAPT and PROMIS-HFpEF studies.

FUNDING SUPPORT AND AUTHOR DISCLOSURES

Mr Jiang is an employee of Us2.ai. Mr Hummel is an employee of Us2.ai. Mr Frost is an employee of Us2.ai. Dr Lam is supported by a Clinician Scientist Award from the National Medical Research Council of Singapore; has received research support from NovoNordisk and Roche Diagnostics; has served as consultant or on the Advisory Board/Steering Committee/Executive Committee for Alleviant Medical, Allysta Pharma, Amgen, AnaCardio AB, Applied Therapeutics, AstraZeneca, Bayer, Boehringer Ingelheim, Boston Scientific, CardioRenal, Cytokinetics, Darma Inc, EchoNous Inc, Eli Lilly, Impulse Dynamics, Intellia Therapeutics, Ionis Pharmaceutical, Janssen Research and Development LLC, Medscape/WebMD Global LLC, Merck, Novartis, Novo Nordisk, Prosciento Inc, Quidel Corporation, Radcliffe Group Ltd, Recardio Inc, ReCor Medical, Roche Diagnostics, Sanofi, Siemens Healthcare Diagnostics and Us2.ai; has served as cofounder and non-executive director of Us2.ai; has a patent pending (PCT/SG2016/050217; Method for diagnosis and prognosis of chronic heart failure); and holds U.S. Patent No. 10,702,247 for automated clinical workflow that recognizes and analyses 2-dimensional and Doppler echo images for cardiac measurements and the diagnosis, prediction, and prognosis of heart disease. Dr Shah is supported by grants from the U.S. National Institutes of Health (National Heart, Lung, and Blood Institute; U54 HL160273, R01 HL140731, and R01 HL149423), AstraZeneca, Corvia Medical, and Pfizer; and has received consulting fees from Abbott, AstraZeneca, Alleviant, Amgen, Aria CV, Axon Therapies, Bayer, Boehringer Ingelheim, Boston Scientific, Bristol-Myers Squibb, Cyclerion, Cytokinetics, Edwards Lifesciences, Eidos, Imara, Impulse Dynamics, Intellia, Ionis, Lilly, Merck, Metabolic Flux, MyoKardia, NGM Biopharmaceuticals, Novartis, Novo Nordisk, Pfizer, Prothena, Regeneron, Rivus, Sardocor, Shifamed, Tenax, Tenaya, Ultromics, and United Therapeutics. Dr Lund is supported by Karolinska Institutet, the Swedish Research Council (grant 523-2014-2336), the Swedish Heart Lung Foundation (grants 20150557, 20190310), and the Stockholm County Council (grants 20170112, 20190525); and unrelated to the present work, Dr Lund has received grants, consulting, and honoraria from Abbot, Alleviant, AstraZeneca, Bayer, Biopeutics, Boehringer Ingelheim, Edwards, Merck/Merck Sharp & Dohme, Novartis, Novo Nordisk, Owkin, Pharmacosmos, Vifor Pharma; and has stock ownership in AnaCardio. Dr Stone has received Speaker or other honoraria from Cook, Terumo, QOOL Therapeutics, and Orchestra Biomed; is a consultant to Valfix, TherOx, Vascular Dynamics, Robocath, HeartFlow, Gore, Ablative Solutions, Miracor, Neovasc, V-Wave, Abiomed, Ancora, MAIA

Pharmaceuticals, Vectorious, Reva, Matrizyme, Cardiomech; and has equity/options from Ancora, Qool Therapeutics, Cagent, Applied Therapeutics, Biostar family of funds, SpectraWave, Orchestra Biomed, Aria, Cardiac Success, MedFocus family of funds, and Valfix. Dr Swaminathan has received consulting fees from US2.ai. Dr Weissman is the Associate Director of an academic core laboratory with research institutional grants/agreements (MedStar Health) with Us2.ai, Ultromics, TOMTEC, GE, Caption Health, egnite, Abbott, Edwards, Medtronic, Boston Scientific, Corcym, Ancora Heart, Neovasc, InnovHeart, and Polares Medical. Dr Asch is the director of an academic core laboratory with research institutional grants/agreements (MedStar Health) with Us2.ai, Ultromics, TOMTEC, GE, Caption Health, egnite, Abbott, Edwards, Medtronic, Boston Scientific, Corcym, Ancora Heart, Neovasc, InnovHeart, Polares Medical, and Foldax; and is on the Scientific Advisory Board (unpaid) for Us2.ai, Ultromics, and Abbott. All other authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Federico M. Asch, Cardiovascular Core Laboratories and Cardiac Imaging Research, MedStar Health Research Institute, 100 Irving Street NW, Suite EB 5123, Washington, DC 20010, USA. E-mail: federico.asch@medstar.net. X handle: @FedericoAsch.

PERSPECTIVES

COMPETENCY IN MEDICAL KNOWLEDGE: Implementing echocardiographic automated ML models for MR severity grading in clinical practice could be critical to screen for patients in need of further clinical evaluation, therapeutic interventions, or for proper referral to specialized valve or HF clinics. The MR grading was associated with 1-year all-cause mortality and HF hospitalizations.

COMPETENCY IN PATIENT CARE AND PROCEDURAL SKILLS: A novel, fully automated, multiparametric ML model for echocardiographic image analysis is highly feasible, fast, and accurate and can, therefore, improve quality and efficiency in the echocardiography laboratory.

TRANSLATIONAL OUTLOOK: The use of novel automated ML models creates a feasible, fast, and accurate workflow for MR severity grading.

REFERENCES

1. Nkomo VT, Gardin JM, Skelton TN, Gottdiener JS, Scott CG, Enriquez-Sarano M. Burden of valvular heart diseases: a population-based study. *Lancet*. 2006;368:1005-1011.

2. Bonow RO, O'Gara PT, Adams DH, et al. 2020 focused update of the 2017 ACC Expert Consensus Decision Pathway on the Management of Mitral Regurgitation. *J Am Coll Cardiol*. 2020;75:2236-2270.

3. Iung B, Delgado V, Rosenhek R, et al. Contemporary presentation and management of valvular heart disease. *Circulation*. 2019;140:1156-1169.

4. Feldman T, Foster E, Glower DD, et al. Percutaneous repair or surgery for mitral regurgitation. *N Engl J Med*. 2011;364:1395-1406.

5. Stone GW, Lindenfeld J, Abraham WT, et al. Transcatheter mitral-valve repair in patients with heart failure. *N Engl J Med*. 2018;379:2307-2318.

6. Lim DS, Smith RL, Gillam LD, et al. Randomized comparison of transcatheter edge-to-edge repair for degenerative mitral regurgitation in prohibitive surgical risk patients. *JACC Cardiovasc Interv*. 2022;15:2523-2536.

7. Zoghbi WA, Adams D, Bonow RO, et al. Recommendations for noninvasive evaluation of native valvular regurgitation: a report from the American Society of Echocardiography developed in collaboration with the Society for Cardiovascular Magnetic Resonance. *J Am Soc Echocardiogr*. 2017;30:303-371.

8. Kar S, Bardeleben RSv, Rottbauer W, et al. Contemporary outcomes following transcatheter edge-to-edge repair. *JACC Cardiovasc Interv*. 2023;16:589-602.

9. Sermesant M, Delingette H, Cochet H, Jaïs P, Ayache N. Applications of artificial intelligence in cardiovascular imaging. *Nat Rev Cardiol*. 2021;18:600-609.

10. Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging. *J Am Coll Cardiol*. 2019;73:1317-1335.

11. Narang A, Bae R, Hong H, et al. Utility of a deep-learning algorithm to guide novices to acquire echocardiograms for limited diagnostic use. *JAMA Cardiol*. 2021;6:624-632.

12. Tromp J, Seekings PJ, Hung CL, et al. Automated interpretation of systolic and diastolic function on the echocardiogram: a multicohort study. *Lancet Digit Health*. 2022;4:e46-e54.

13. Tromp J, Bauer D, Claggett BL, et al. A formal validation of a deep learning-based automated workflow for the interpretation of the echocardiogram. *Nat Commun*. 2022;13:6776.

14. Ouyang D, He B, Ghorbani A, et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature*. 2020;580:252-256.

15. Asch FM, Poilvert N, Abraham T, et al. Automated echocardiographic quantification of left ventricular ejection fraction without volume measurements using a machine learning algorithm mimicking a human expert. *Circulation Cardiovas Imag*. 2019;12:e009303.

16. Asch FM, Descamps T, Sarwar R, et al. Human versus artificial intelligence-based echocardiographic analysis as a predictor of outcomes: an analysis from the World Alliance Societies of Echocardiography COVID Study. *J Am Soc Echocardiogr*. 2022;35:1226-1237.e7.

17. Madani A, Arnaout R, Mofrad M, Arnaout R. Fast and accurate view classification of echocardiograms using deep learning. *NPJ Digital Med*. 2018;1:6.

18. Lang RM, Addetia K, Miyoshi T, et al. Use of machine learning to improve echocardiographic image interpretation workflow: a disruptive paradigm change? *J Am Soc Echocardiogr*. 2021;34:443-445.

19. Otto CM, Nishimura RA, Bonow RO, et al. 2020 ACC/AHA Guideline for the Management of Patients With Valvular Heart Disease. *J Am Coll Cardiol*. 2021;77:e25-e197.

20. Vahanian A, Beyersdorf F, Praz F, et al. 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). *Eur Heart J*. 2021;43:561-632.

21. Shah SJ, Lam CSP, Svedlund S, et al. Prevalence and correlates of coronary microvascular dysfunction in heart failure with preserved ejection fraction: PROMIS-HFpEF. *Eur Heart J*. 2018;39:3439-3450.

22. Tseng AS, Lopez-Jimenez F, Pellikka PA. Future guidelines for artificial intelligence in echocardiography. *J Am Soc Echocardiogr*. 2022;35:878-882.

23. Castello R, Lenzen P, Aguirre F, Labovitz A. Variability in the quantitation of mitral regurgitation by Doppler color flow mapping: comparison of transthoracic and transesophageal studies. *J Am Coll Cardiol*. 1992;20:433-438.

24. Dabiri Y, Yao J, Mahadevan VS, et al. Mitral valve atlas for artificial intelligence predictions of MitraClip intervention outcomes. *Front Cardiovasc Med*. 2021;8:759675.

25. Rashmi N, Bo W, Wendy T. Artificial intelligence for the echocardiographic assessment of valvular heart disease. *Heart*. 2022;108:1592-1599.

26. Yang F, Chen X, Lin X, et al. Automated analysis of doppler echocardiographic videos as a screening tool for valvular heart diseases. *JACC Cardiovasc Imaging*. 2022;15:551-563.

27. Moghaddasi H, Nourian S. Automatic assessment of mitral regurgitation severity based on extensive textural features on 2D echocardiography videos. *Comput Biol Med*. 2016;73:47-55.

28. Edwards LA, Feng F, Iqbal M, et al. Machine learning for pediatric echocardiographic mitral regurgitation detection. *J Am Soc Echocardiogr*. 2023;36:96-104.e4.

29. Dziadzko V, Clavel M-A, Dziadzko M, et al. Outcome and undertreatment of mitral regurgitation: a community cohort study. *Lancet*. 2018;391:960-969.

30. Nishimura RA, Otto CM, Bonow RO, et al. 2014 AHA/ACC guideline for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol*. 2014;63:e57-e185.

KEY WORDS artificial intelligence, continuous wave Doppler density, echocardiography, machine learning, mitral regurgitation

APPENDIX For supplemental tables and figures, please see the online version of this paper.