

Deep learning-based prediction of right ventricular ejection fraction using 2D echocardiograms

Márton Tokodi, M.D., Ph.D.^{1*}, Bálint Magyar, M.Sc.^{2*}, András Soós, M.Sc.^{1,2}, Masaaki Takeuchi, M.D., Ph.D.³, Máté Tolvaj¹, Bálint Károly Lakatos, M.D., Ph.D.¹, Tetsuji Kitano, M.D., Ph.D.⁴, Yosuke Nabeshima, M.D., Ph.D.⁵, Alexandra Fábián, M.D.¹, Mark Bence Szigeti, B.Sc.², András Horváth, M.Sc., Ph.D.^{2#}, Béla Merkely, M.D., Ph.D., D.Sc.^{1#}, Attila Kovács, M.D., PhD.^{1#}

*Márton Tokodi, M.D., Ph.D., and Bálint Magyar, M.Sc. contributed equally to this work and are joint first authors.

#András Horváth, M.Sc., Ph.D., Béla Merkely, M.D., Ph.D., D.Sc., and Attila Kovács, M.D., Ph.D. contributed equally to this work and are joint last authors.

¹Heart and Vascular Center, Semmelweis University, Budapest, Hungary

²Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary

³Department of Laboratory and Transfusion Medicine, University of Occupational and Environmental Health Hospital, Kitakyushu, Japan

⁴Department of Cardiology and Nephrology, Wakamatsu Hospital of the University of Occupational and Environmental Health, Kitakyushu, Japan

⁵Second Department of Internal Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan

Brief Title: DL-based prediction of RVEF using 2D echocardiograms

SOURCES OF FUNDING

This study was supported by the Thematic Excellence Program (2020-4.1.1. – TKP2020) of the Ministry for Innovation and Technology in Hungary, within the framework of the Bioimaging thematic program of the Semmelweis University. This project was also supported

by a grant from the National Research, Development, and Innovation Office (NKFIH) of Hungary (FK 142573 to Attila Kovács). Project no. RRF-2.3.1-21-2022-00004 (MILAB) has been implemented with the support provided by the European Union.

CONFLICT OF INTEREST

Márton Tokodi, Bálint Magyar, Bálint Károly Lakatos, Alexandra Fábián, and Attila Kovács report personal fees from Argus Cognitive, Inc., outside the submitted work. András Soós reports personal fees from Quantib B.V., outside the submitted work. Masaaki Takeuchi reports grants from Philips Medical Systems, outside the submitted work. Béla Merkely reports grants from Boston Scientific and Medtronic and personal fees from Biotronic, Abbott, Astra Zeneca, Novartis, and Boehringer-Ingelheim, outside the submitted work. All other authors have no conflict of interest to declare.

CORRESPONDENCE

Márton Tokodi, M.D., Ph.D.

Heart and Vascular Center, Semmelweis University

Address: 68 Városmajor Street, Budapest 1122, Hungary

Phone: +3630-465-23-56

Fax: +361-458-68-42

E-mail: tokmarton@gmail.com

Attila Kovács, M.D., Ph.D.

Heart and Vascular Center, Semmelweis University

Address: 68 Városmajor Street, Budapest 1122, Hungary

Phone: +3620-666-34-27

Fax: +361-458-68-42

E-mail: attila.kovacs@med.semmelweis-univ.hu

ABSTRACT

Background: Evidence has shown the independent prognostic value of right ventricular (RV) function, even in patients with left-sided heart disease. The most widely used imaging technique to measure RV function is echocardiography; however, conventional two-dimensional (2D) echocardiographic assessment is unable to leverage the same clinical information that three-dimensional (3D) echocardiography-derived RV ejection fraction (RVEF) can provide.

Objective: We aimed to implement a deep learning (DL)-based tool to estimate RVEF from 2D echocardiographic videos. In addition, we benchmarked the performance of our system against human expert reading and evaluated the prognostic power of the predicted RVEF values.

Methods: We retrospectively identified 831 patients with RVEF measured by 3D echocardiography. All 2D apical 4-chamber view echocardiographic videos of these patients were retrieved (n=3,583), and each subject was assigned to either the training or the internal validation set (80:20 ratio). Using the videos, several spatiotemporal convolutional neural networks were trained to predict RVEF. The three best-performing networks were combined into an ensemble model, which was further evaluated in an external dataset containing 1,493 videos of 365 patients with a median follow-up time of 1.9 years.

Results: Our model predicted RVEF with a mean absolute error of 4.56 percentage points in the internal and 5.54 percentage points in the external validation set. In the latter, the model identified RV dysfunction (defined as RVEF <45%) with an accuracy of 78.4%, which was comparable to an expert reader's visual assessment (77.0%, McNemar's test $p=0.678$). The DL-predicted RVEF values were associated with major adverse cardiac events independent of age, sex, and left ventricular systolic function (hazard ratio: 0.924 [95% confidence interval: 0.862-0.990], $p=0.025$).

Conclusions: Using 2D echocardiographic videos alone, our DL-based tool can accurately assess RV function, with similar diagnostic and prognostic power as 3D imaging.

Keywords: echocardiography, deep learning, right ventricle, right ventricular ejection fraction, right ventricular dysfunction

ABBREVIATIONS

2D – two-dimensional

3D – three-dimensional

AUC – area under the receiver operating characteristic curve

CI – confidence interval

CMR – cardiac magnetic resonance

DL – deep learning

HR – hazard ratio

LV – left ventricular

LVEF – left ventricular ejection fraction

MACE – major adverse cardiac event

MAE – mean absolute error

RV – right ventricular

RVEF – right ventricular ejection fraction

INTRODUCTION

The prognostic significance of right ventricular (RV) dysfunction has been well recognized in various cardiovascular conditions (1). It not only represents the major determinant of clinical symptoms and survival in patients with pulmonary hypertension (2) but also has an independent prognostic value in left-sided heart diseases (3). About half of the patients with heart failure and reduced left ventricular ejection fraction (LVEF) suffer from RV dysfunction, which is associated with a 2-fold increased risk of adverse outcomes (4). Therefore, screening for RV systolic dysfunction should be an integral part of cardiovascular care.

Two-dimensional (2D) echocardiography is the most frequently performed test to assess RV function (5-7). However, conventional 2D parameters are unable to reliably capture RV dysfunction across the entire spectrum of cardiac diseases (6,7). Three-dimensional (3D) echocardiography-derived RV ejection fraction (RVEF) offers a panacea for their shortcomings as it is a sensitive and reproducible parameter that has been validated against gold-standard cardiac magnetic resonance (CMR) imaging (7,8). Nonetheless, 3D echocardiography has limited availability, is more time-consuming, and requires significant human expertise. Therefore, we constantly pursue novel automated tools that utilize readily available and routinely acquired echocardiographic recordings to detect RV dysfunction.

In recent years, deep learning (DL) has enabled automated systems to meet or even surpass the performance of clinical experts across a range of image analysis tasks in cardiovascular medicine, from diagnosis to prognostication (9,10). One of the potential areas of application where DL has been successfully deployed is the automated, real-time measurement of LVEF using conventional echocardiographic acquisitions (11-14). On the other hand, no DL-based tool has been proposed yet that infers RVEF directly from a single 2D echocardiographic video which might be attributable to the more complex shape and

function of the RV.

Accordingly, we aimed to implement a DL-based tool for estimating RVEF from 2D echocardiographic videos. In addition, we benchmarked the performance of our system against human expert reading and evaluated the prognostic power of the predicted RVEF values.

METHODS

Overview of the study protocol

Our study revolved around constructing a fully automated DL pipeline that analyzes a single 2D apical 4-chamber view echocardiographic video to predict 3D echocardiography-derived RVEF as a continuous variable (Central Illustration). To ensure the robustness and generalizability of the DL model, we included two patient cohorts from two separate centers. A large single-center cohort (i.e., internal dataset) was used for model training and internal validation, whereas a second smaller cohort from another center served as a dataset for external validation. In the latter, we also benchmarked the model's performance against an expert reader's visual assessment and explored the association of the predicted RVEF values with clinical outcomes. Although we trained our DL model to perform regression, we also evaluated the model's capability to discern patients with RV dysfunction (3D echocardiography-derived RVEF <45%) from those with normal RV function (3D echocardiography-derived RVEF $\geq$ 45%). For each component of the study protocol, a comprehensive description is provided in the following subsections or the Supplementary Appendix.

The study protocol conforms with the principles outlined in the Declaration of Helsinki, and it was approved by the Semmelweis University Regional and Institutional Committee of Science and Research Ethics (approval No. 190/2020) and the ethics committee of the University of Occupational and Environmental Health (approval No.

UOEHCRB21-144). Obtaining informed consent was waived owing to the retrospective nature of the analysis.

Internal dataset

By retrospectively identifying 3D transthoracic echocardiographic examinations performed between November 2013 and March 2021 at the Heart and Vascular Center of Semmelweis University (Budapest, Hungary), we created a comprehensive database comprising a wide range of patients, including individuals with cardiovascular disorders, healthy sedentary volunteers, and elite athletes. The inclusion criteria were the following: (1) at least one available apical 4-chamber view (either standard or RV-focused) video, and (2) 3D analysis and calculation of RVEF feasible on the acquired 3D datasets. Videos with (1) invalid heart rate or frame per second values in the DICOM tags, (2) acquisition issues comprising but not limited to severe translational motion, gain changes, depth changes, view changes, sector position changes, and (3) an inadequate number of frames (i.e., videos shorter than one cardiac cycle or containing <20 frames per cardiac cycle) were discarded. For each patient, clinical characteristics and medical history were retrieved from electronic medical records.

External dataset

In a second center (Hospital of the University of Occupational and Environmental Health, Kitakyushu, Japan), we retrospectively identified patients who had undergone 3D transthoracic echocardiographic examinations between January 2014 and December 2020. As the majority of this patient cohort also underwent CMR imaging examination on the same day (details of the CMR imaging protocol are provided in the Supplementary Appendix), patient selection was related to the clinical indications for CMR. Inclusion criteria were the same as in the internal dataset, with the additional requirement that outcomes data were available.

Patient characteristics, medical history, and outcome data were retrieved for all patients from the electronic medical records system of the external center in November 2021. The attending physician or next of kin was contacted if a patient had been treated at a different center. The primary endpoint of our study was the occurrence of a major adverse cardiac event (MACE, defined as the composite of cardiac death, worsening of heart failure requiring hospitalization, or ventricular tachycardia or fibrillation).

Echocardiographic protocol

Each patient in both cohorts underwent transthoracic echocardiographic examination by experienced echocardiographers using commercially available ultrasound scanners (Vivid E95 system, GE Vingmed Ultrasound, Horten, Norway; iE33, EPIQ CVx, 7C, or 7G systems, Philips, Best, The Netherlands). Beyond the conventional echocardiographic protocol, electrocardiogram-gated full-volume 3D datasets reconstructed from four cardiac cycles were also acquired from an RV-optimized apical view, targeting a minimum volume rate of 25 volumes/second. 3D datasets were analyzed, and RVEF was computed using a commercially available software solution (4D RV-Function 2, TomTec Imaging, Unterschleissheim, Germany) on a separate workstation. To assess the inter-reader variability of RVEF, a second experienced operator (blinded to the measurements of the first operator) analyzed the 3D recordings of 35 randomly selected patients from the external validation cohort. LV morphology and function were also assessed using 3D echocardiography (4D LV-Analysis 3, TomTec Imaging, Unterschleissheim, Germany).

A single experienced echocardiographer reviewed all videos and images in each patient's echocardiographic examination and identified 2D apical 4-chamber view loops (either standard or RV-focused, no color Doppler applied). These videos were then exported as DICOM files and were used as the input for the DL models.

Visual assessment of image quality and right ventricular dysfunction

An experienced echocardiographer reviewed all 2D echocardiographic videos of the internal and the external datasets to (1) assess the image quality using a 5-point Likert scale (non-diagnostic, poor, moderate, good, excellent), (2) determine the view type (standard or RV-focused), and (3) assess video inclusion and exclusion criteria. Another expert reader (with more than 10 years of experience in 3D echocardiography-based RV assessment) from the first center was tasked to review the videos of the external dataset and to identify the presence of RV dysfunction (defined as RVEF <45%) by simple visual estimation, blinded to the rest of the echocardiographic study, the diagnosis, and the clinical characteristics of the patients.

Model training and evaluation

After splitting each video into image sequences containing frames of one cardiac cycle, we performed a multi-step preprocessing, which is described thoroughly in the Supplementary Appendix. Then, we implemented several spatiotemporal neural networks to predict RVEF directly from the preprocessed image sequences and the corresponding binary masks (Supplementary Figure 5). Despite the availability of CMR-derived RVEF in the external dataset, 3D echocardiography-derived RVEF was considered the ground truth in both the internal and external datasets. Each network consists of three key components: a feature extractor (ResNext (15) or ShuffleNet (16)) that derives unique features from each frame, a temporal convolutional layer that combines these pieces of information to evaluate temporal changes in the image sequence, and a regression head (comprising two fully connected layers) which returns a single predicted RVEF value. The optimal set of hyperparameters was selected using grid search (Supplementary Table 2). To enable the internal validation of the DL models, the internal dataset was split in an approximately 80:20 ratio into train and internal validation sets. Splitting was performed at the patient level to avoid data leakage (i.e., we assigned all videos of a given patient either to the train or the

internal validation set). Dataset balancing (in combination with augmentation techniques) was applied on the train set to improve generalization. The models' performance was evaluated using the mean absolute error (MAE) and the coefficient of determination (R^2). The three best-performing networks were combined into an ensemble model (Supplementary Table 3), which we also evaluated in the external dataset. As models were trained to analyze one cardiac cycle at a time, the average of these predictions was calculated to get a single predicted RVEF value for a given video. Besides assessing the model's performance at the video level, we also calculated the evaluation metrics at the examination level using the mean values of the video-level predictions (i.e., by averaging the predicted RVEF values of all videos acquired for a given patient during the same examination).

As a final step, we applied occlusion sensitivity visualization to generate saliency maps providing insight into the prediction process (17). This technique systematically occludes small segments of the input image to measure the impact of the given region on the model's performance. If the network's performance decreases compared to the baseline results, the occluded region contains relevant anatomical structures. If the performance improves, the occluded region contains information that confuses the model. As our model processed image sequences consisting of 20 frames, we occluded one patch only in one image at a time. By denoting important regions with green and confusing regions with red color, a heatmap is created that can be superimposed on the original images to build saliency maps. Then, we can inspect these maps to ascertain that predictions are made based on highly relevant anatomical structures (e.g., the RV myocardium) while other less relevant regions (e.g., extracardiac structures) and potential confounders (e.g., remnants of the electrocardiogram signal) are neglected. The saliency maps presented in the manuscript were generated by applying patches of 8x8 pixels as they provided a better resolution-interpretability trade-off than the application of smaller (4x4 pixels) or bigger (16x16 and

32x32 pixels) patches.

A detailed description of preprocessing, model training and evaluation, visualization, and statistical analysis is provided in the Supplementary Appendix. Methods and results are reported in compliance with the Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME) checklist (Supplementary Table 4) (18).

RESULTS

Patient and video selection

From 3,707 apical 4-chamber view echocardiographic videos of 840 subjects fulfilling the inclusion criteria in the first center, we discarded 9 (0.24%) videos due to invalid heart rate or frame rate values in the DICOM tags, 109 (2.94%) videos due to acquisition issues (e.g., severe translational motion, gain changes, depth changes, view changes, sector position changes), and 6 (0.16%) videos that had an inadequate number of frames for DL analysis (Figure 1). The final internal dataset contained 3,583 echocardiographic videos from 944 echocardiographic examinations of 831 individuals. To promote further innovation, we made the internal dataset publicly available (RVENet dataset) (19).

In the external center, 1,536 apical 4-chamber view echocardiographic videos of 371 patients were identified retrospectively (Figure 1). We excluded 5 (0.33%) videos due to invalid heart rate or frame rate values in the DICOM tags, 31 (2.02%) videos due to acquisition issues, and 7 (0.46%) videos that had an inadequate number of frames for analysis. The final external dataset included 1,493 echocardiographic videos from 365 examinations of 365 patients.

Clinical characteristics of the study cohorts

The internal dataset comprised seven distinct cohorts: (1) healthy volunteers (n=192, without history and risk factors of cardiovascular diseases and good LV and RV function);

(2) elite, competitive athletes (n=139); (3) cardiomyopathy patients (n=125, 40.0% of ischemic origin, median LVEF of 30.1%); (4) valvular heart disease patients (n=155, 54.8% with severe aortic valve disease, 43.9% with severe mitral valve disease, 1.3% who previously underwent surgical mitral valve replacement or repair); (5) clinically stable patients who previously underwent orthotopic heart transplantation (n=87); (6) pediatric subjects (n=77, healthy volunteers and patients who previously underwent renal transplantation); and (7) others (n=56). The clinical and echocardiographic characteristics of this dataset are described comprehensively in Supplementary Table 5. RVEF values ranged from 21.6 to 75.5%, with a median of 54.8% (Supplementary Figure 6). RV dysfunction (defined as RVEF <45%) was identified in 135 (14.3%) examinations (containing 523 videos).

The external dataset included only adult patients with various cardiac disorders (e.g., cardiomyopathies, valvular heart diseases, pulmonary hypertension). The majority of these patients (97.0%) also underwent CMR imaging examination on the same day. During the median follow-up time of 1.9 years, 49 patients (13.4%) reached the primary endpoint (12 cardiac death, 31 worsening of heart failure requiring hospitalization, 6 ventricular tachycardia or fibrillation). Patients with MACE were older, had worse functional status, and showed a higher prevalence of chronic kidney disease than those with no events during the follow-up period (Table 1). LV and RV morphological and functional parameters were more impaired in patients who experienced MACE (Table 1). RVEF values ranged from 21.9 to 65.2%, with a median of 47.6% (Supplementary Figure 7). RV dysfunction (i.e., RVEF <45%) was diagnosed in 145 (39.7%) patients (632 videos).

Prediction of RVEF

At the examination level, our DL model predicted RVEF with an MAE of 4.57 percentage points ($R^2=0.52$) in the internal and 5.54 percentage points ($R^2=0.45$) in the

external validation set. At the video level, the model achieved an MAE of 5.06 percentage points ($R^2=0.50$) and 6.11 percentage points ($R^2=0.33$), respectively (Figure 2, Supplementary Table 6). Our model performed similarly well in patients with convergent and divergent LVEF and RVEF (Supplementary Table 7), as well as in healthy subjects and patients with cardiovascular disease (Supplementary Table 8). Bland-Altman analysis showed a statistically significant but small bias between the DL-predicted and the ground truth RVEF values in the internal dataset (per-video bias: -0.698 percentage points, $p=0.007$; per-examination bias: -0.944 percentage points, $p=0.044$) but no significant bias in the external dataset (per-video bias: 0.264 percentage points, $p=0.590$; per-examination bias: 0.242 percentage points, $p=0.921$; Supplementary Figure 8). The performance of our model is even more impressive in light of the inter-reader variability of RVEF, which was 4.30 percentage points.

We also explored whether image quality, the type of apical 4-chamber view (standard vs. RV-focused), or the utilization of different ultrasound systems impact the magnitude of prediction errors. Despite showing a slight tendency to predict RVEF with smaller errors from videos with good and excellent quality (Supplementary Figure 9 and Supplementary Table 9), our model did not exhibit a considerable variation in performance across the image quality categories (internal validation set: Kruskal-Wallis test $p=0.340$, external validation set: Kruskal-Wallis test $p=0.309$). Moreover, there was no difference in prediction errors between standard and RV-focused apical 4-chamber view videos either in the internal (Mann-Whitney U test $p=0.493$) or the external dataset (Mann-Whitney U test $p=0.096$, Supplementary Figure 9 and Supplementary Table 10). In the internal validation set, the magnitude of prediction errors was similar across the subsets of videos acquired using different ultrasound systems (Kruskal-Wallis test $p=0.185$), but we could observe a few minor differences in the external validation set (Kruskal Wallis test $p=0.025$, Supplementary

Figure 10 and Supplementary Table 11). Still, the MAE remained within acceptable limits even in the worst-performing subset of videos (i.e., those acquired using the Philips iE33 ultrasound system).

Detection of RV dysfunction

We used the guideline-directed cut-off value of 45% by 3D echocardiography to define RV dysfunction. We were able to detect RV dysfunction based on the DL-derived RVEF values with high discriminatory power in both the internal (per-video area under the receiver operating characteristic curve [AUC]: 0.929 [95% confidence interval [CI]: 0.896-0.957], per-examination AUC: 0.915 [95% CI: 0.852-0.966]) and the external datasets (per-video AUC: 0.807 [95% CI: 0.785-0.829], per-examination AUC: 0.841 [95% CI: 0.796-0.883]) (Figure 3 and Supplementary Table 12). In addition, both the per-video and per-examination accuracy of the DL model was comparable to that of the expert human reader in the external dataset (per-video accuracy: 74.3 vs. 75.6%, $p=0.389$, per-examination accuracy: 78.4 vs. 77.0%, $p=0.678$, Supplementary Table 13), whereas it exhibited higher sensitivity (per-video sensitivity: 67.7 vs. 50.3%, $p<0.001$, per-examination sensitivity: 67.6 vs. 51.7%, $p<0.001$, Supplementary Table 13) at the expense of lower specificity (per-video specificity: 79.1 vs. 94.1%, $p=0.003$, per-examination specificity: 85.5 vs. 93.6%, $p=0.006$, Supplementary Table 13).

Association with outcomes

Predictors of MACE in the external cohort are listed in Supplementary Table 14. Similar to the ground truth RVEF (hazard ratio [HR]: 0.936 [95% CI: 0.911-0.962], $p<0.001$), the DL-predicted values were also associated with MACE (HR: 0.912 [95% CI: 0.875-0.950], $p<0.001$). In a multivariable Cox model including age, sex, and LVEF as covariates, the DL-predicted RVEF was found to be an independent predictor of MACE (HR: 0.924 [95% CI: 0.862-0.990], $p=0.025$, Table 2).

Using the standard cut-off value of 45%, DL-predicted RVEF was able to discriminate between a high-risk and a low-risk group in terms of MACE (Figure 4). Based on the value of predicted RVEF, patients were also split into four quartiles. As depicted by Kaplan–Meier curves, there was a significant difference in the occurrence of events across the quartiles, and a graded increase could be observed in event rates (Figure 4). These results were similar concerning the ground truth 3D echocardiography-derived values.

Saliency maps and visualization

By denoting important regions with green and confusing regions with red color, saliency maps were created and superimposed on the original images to gain insight into the prediction process (Supplementary Videos 1, 2, and 3). In Figure 5, we depicted multiple color-coded systolic and diastolic frames from representative videos of three patients. We could observe that the specific regions of the heart contributed to the predictions of RVEF at different phases of the cardiac cycle.

DISCUSSION

In this study, we proposed a DL-based tool capable of accurately estimating RVEF from 2D echocardiographic videos. In addition, our system enabled us to identify RV dysfunction with a performance similar to an expert human reader. Moreover, not only the diagnostic but also the prognostic potential of our system was demonstrated as impaired values of DL-predicted RVEF were associated with a significantly higher risk of MACE, independent of age, sex, and LV function.

2D echocardiography is the first-of-choice non-invasive imaging technique to assess the morphology and function of the RV (5-7). However, the complex geometry and mechanical pattern of the chamber hinder the confident detection of RV dysfunction even if a multiparametric 2D echocardiography-based approach is applied (20,21). The most commonly measured parameters have several shortcomings (e.g., overestimation or

underestimation of global systolic function, load dependency, or technical limitations), which should be considered while utilizing them to evaluate RV function in different pathological scenarios (6,7). Many of these limitations can be circumvented by 3D echocardiography and CMR as they offer a unique opportunity to map the entire endocardial surface of the RV independent of a priori assumptions about its shape and function to provide a diagnostic and prognostic functional parameter – RVEF. Still, 3D echocardiography is underutilized in the clinical routine and is primarily available in academic centers (22). Similarly, CMR imaging is not widely accessible due to its high costs and significant dependence on clinical expertise (23,24).

The simultaneous interpretation of multiple echocardiographic features that can be suggestive of RV dysfunction is a challenging task. Therefore, the potential influence of complex but often hidden hemodynamic and mechanical interactions is overlooked. By digesting the abundance of meaningful information encompassed by stacks of 2D echocardiographic images, DL models may be capable of efficiently integrating all these features to gain knowledge of the same value as 3D echocardiography-derived RVEF. Lately, several DL-based solutions have been proposed for the automated assessment of LVEF from 2D echocardiograms (11-14). Importantly, most of these systems rely on supervised segmentation algorithms to outline cardiac chambers. Even though similar segmentation-based approaches might be utilized to measure RV fractional area change from either 2D echocardiographic recordings or CMR images (25,26), a different concept would be required to infer RVEF from single-plane 2D image sequences. As information enabling the direct calculation of RVEF is not available in such videos, we had to design a system capable of extracting also features that are indirect indicators of RV function.

Our tool achieved an MAE of 4.57 and 5.54 percentage points in the internal and external datasets at the examination level, respectively. This remarkable performance might

be attributable to the fact that the system learned to extract spatiotemporal information not only from the RV myocardium but also from other cardiac structures such as the LV, atria, and valves. This phenomenon resonates well with the physiological concepts of ventricular interdependence and atrioventricular coupling that highlight the significance of high-level interactions and synchrony between the different cardiac structures. By extracting these relevant pieces of information, our DL model can accurately predict RVEF – the most integrative parameter of RV systolic function. Of note, the model’s performance was only minimally affected by the image quality, the view plane, or the ultrasound system used.

RVEF of 45% is a meaningful and standard cut-off value to diagnose RV dysfunction, as lower values are clearly associated with adverse outcomes (27). Notably, the discriminative power of our DL-based tool was comparable to human expert reading in distinguishing patients with RV dysfunction from those with normal RV function. By exhibiting a significantly higher sensitivity (at the expense of a lower specificity) than an experienced cardiologist, our tool is even more appealing and may be used for screening purposes. Beyond accurately predicting the exact value of RVEF and the presence of RV dysfunction, it is essential to emphasize that the predictions preserved the prognostic potentials of the ground truth 3D echocardiography-based measurements to identify distinct subgroups of patients exhibiting increased risk of MACE. Importantly, the prognostic value was independent of age, sex, and LV systolic function. Although our system, to our knowledge, is the first to estimate RVEF based on 2D echocardiographic videos, Shad et al. have recently deployed a segmentation-free DL approach to predict the onset of postoperative RV failure in patients undergoing LV assist device implantation (28). Despite the apparent differences in study design, preprocessing, and model development, they also reported that DL could efficiently forecast future adverse events, which is in line with our current findings.

Our system makes predictions from an image sequence of one cardiac cycle within a

few hundred milliseconds on a single Nvidia GeForce RTX 2080 graphics processing unit. Therefore, it may operate on echocardiographic machines in real time or as a cloud-based service. We foresee that similar tools will provide reasonable alternatives to more advanced but less available modalities and complex parameters as they may harvest the same diagnostic and prognostic information from lower-level imaging.

While our DL-based tool was trained using the largest 3D echocardiographic dataset of its kind, there remain several limitations of our work. First, the retrospective design and the inclusion of subjects only with 3D echocardiographic datasets suitable for RV analysis may represent a significant source of selection bias. Second, it should be noted that the internal dataset was derived from a single-center echocardiography database. Therefore, it represents the patient population treated at the Heart and Vascular Center of Semmelweis University, and some patient subgroups might be overrepresented, whereas other subgroups (e.g., patients with pulmonary arterial hypertension) might be precluded from our study. Third, there are apparent differences in the demographics and clinical characteristics of the internal and external datasets. We were well aware of these differences, and actually, our study was intentionally designed this way. For model training, we wanted to use a large cohort of patients exhibiting a broad spectrum of RV morphology and function to create a model as generalizable as possible. Then, we evaluated our model on a holdout set of this cohort (i.e., the internal validation set) to see how it performs in patients from the same center with similar clinical and echocardiographic characteristics. As a final step, we wanted to validate the model in an adult patient cohort from an external center with mixed disease etiologies and a higher prevalence of RV dysfunction, as this allowed us to glimpse into the model's utility in a real-life clinical scenario. Thus, given these differences, we expected to observe slightly lower performance in the external dataset than in the internal dataset, which is a common phenomenon in machine learning studies. Nevertheless, our model's

performance was still acceptable in the external dataset. Still, our model should be evaluated in additional patient cohorts exhibiting specific RV remodeling patterns, such as in patients with RV pressure or volume overload or congenital heart diseases. Last, we would like to highlight that an operating range (21.6 – 75.5%) had to be defined based on the ground truth RVEF values of the train set. As machine learning models usually perform poorly if forced to extrapolate, the utilization of our tool on echocardiographic videos of patients with RVEF values outside the operating range should be avoided.

CONCLUSIONS

Using 2D echocardiographic videos, our DL-based tool could accurately assess RV function and leverage similar diagnostic and prognostic information as 3D imaging. Our findings prove that artificial intelligence can gain advanced clinical knowledge from lower-level imaging modalities with a performance that may match or even surpass clinical experts and, therefore, enables the development of accurate, widely accessible, and cost-effective tools in medical imaging.

CLINICAL PERSPECTIVES

Competency in Medical Knowledge:

We implemented a fully automated DL-based tool that could accurately predict RVEF from 2D apical 4-chamber view echocardiographic videos. In addition, it could also identify RV dysfunction with a performance similar to an expert human reader, and it provided the same prognostic information as 3D echocardiography.

Translational Outlook:

In scenarios where 3D echocardiography is unavailable, our DL-based tool may offer an enticing alternative for RV function assessment and may be applied for screening for RV dysfunction. In addition, as a part of point-of-care examinations, it may be used for the real-time and bedside evaluation of RV function, supporting clinical decision-making.

REFERENCES

1. Konstam MA, Kiernan MS, Bernstein D et al. Evaluation and Management of Right-Sided Heart Failure: A Scientific Statement From the American Heart Association. *Circulation* 2018;137:e578-e622.
2. Hassoun PM. Pulmonary Arterial Hypertension. *New England Journal of Medicine* 2021;385:2361-2376.
3. Surkova E, Muraru D, Genovese D, Aruta P, Palermo C, Badano LP. Relative Prognostic Importance of Left and Right Ventricular Ejection Fraction in Patients With Cardiac Diseases. *J Am Soc Echocardiogr* 2019;32:1407-1415.e3.
4. Yancy CW, Jessup M, Bozkurt B et al. 2013 ACCF/AHA Guideline for the Management of Heart Failure. *Journal of the American College of Cardiology* 2013;62:e147-e239.
5. Zaidi A, Knight DS, Augustine DX et al. Echocardiographic assessment of the right heart in adults: a practical guideline from the British Society of Echocardiography. *Echo Res Pract* 2020;7:G19-g41.
6. Rudski LG, Lai WW, Afilalo J et al. Guidelines for the Echocardiographic Assessment of the Right Heart in Adults: A Report from the American Society of Echocardiography: Endorsed by the European Association of Echocardiography, a registered branch of the European Society of Cardiology, and the Canadian Society of Echocardiography. *Journal of the American Society of Echocardiography* 2010;23:685-713.
7. Lang RM, Badano LP, Mor-Avi V et al. Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr* 2015;28:1-39.e14.

8. Muraru D, Spadotto V, Cecchetto A et al. New speckle-tracking algorithm for right ventricular volume analysis from three-dimensional echocardiographic data sets: validation with cardiac magnetic resonance and comparison with the previous analysis tool. *Eur Heart J Cardiovasc Imaging* 2016;17:1279-1289.
9. Litjens G, Ciompi F, Wolterink JM et al. State-of-the-Art Deep Learning in Cardiovascular Image Analysis. *JACC Cardiovasc Imaging* 2019;12:1549-1565.
10. Kagiya N, Tokodi M, Sengupta PP. Machine Learning in Cardiovascular Imaging. *Heart Fail Clin* 2022;18:245-258.
11. Ouyang D, He B, Ghorbani A et al. Video-based AI for beat-to-beat assessment of cardiac function. *Nature* 2020;580:252-256.
12. Asch FM, Mor-Avi V, Rubenson D et al. Deep Learning-Based Automated Echocardiographic Quantification of Left Ventricular Ejection Fraction: A Point-of-Care Solution. *Circulation: Cardiovascular Imaging* 2021;14:e012293.
13. Zhang J, Gajjala S, Agrawal P et al. Fully Automated Echocardiogram Interpretation in Clinical Practice. *Circulation* 2018;138:1623-1635.
14. Liu X, Fan Y, Li S et al. Deep learning-based automated left ventricular ejection fraction assessment using 2-D echocardiography. *Am J Physiol Heart Circ Physiol* 2021;321:H390-h399.
15. Xie S, Girshick R, Dollár P, Tu Z, He K. Aggregated Residual Transformations for Deep Neural Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017:5987-5995.
16. Ma N, Zhang X, Zheng H-T, Sun J. ShuffleNet V2: Practical Guidelines for Efficient CNN Architecture Design. Cham: Springer International Publishing, 2018:122-138.
17. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. Cham: Springer International Publishing, 2014:818-833.

18. Sengupta PP, Shrestha S, Berthon B et al. Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME): A Checklist: Reviewed by the American College of Cardiology Healthcare Innovation Council. JACC: Cardiovascular Imaging 2020;13:2017-2035.
19. <https://rvenet.github.io/dataset/>.
20. Kovács A, Lakatos B, Tokodi M, Merkely B. Right ventricular mechanical pattern in health and disease: beyond longitudinal shortening. Heart Fail Rev 2019;24:511-520.
21. Sanz J, Sánchez-Quintana D, Bossone E, Bogaard HJ, Naeije R. Anatomy, Function, and Dysfunction of the Right Ventricle: JACC State-of-the-Art Review. Journal of the American College of Cardiology 2019;73:1463-1482.
22. Schneider M, Aschauer S, Mascherbauer J et al. Echocardiographic assessment of right ventricular function: current clinical practice. Int J Cardiovasc Imaging 2019;35:49-56.
23. Goldfarb JW, Weber J. Trends in Cardiovascular MRI and CT in the U.S. Medicare Population from 2012 to 2017. Radiol Cardiothorac Imaging 2021;3:e200112.
24. Keenan NG, Captur G, McCann GP et al. Regional variation in cardiovascular magnetic resonance service delivery across the UK. Heart 2021;107:1974-1979.
25. Arai H, Kawakubo M, Sanui K, Iwamoto R, Nishimura H, Kadokami T. Assessment of Bi-Ventricular and Bi-Atrial Areas Using Four-Chamber Cine Cardiovascular Magnetic Resonance Imaging: Fully Automated Segmentation with a U-Net Convolutional Neural Network. International Journal of Environmental Research and Public Health 2022;19:1401.
26. Leng S, Yang X, Zhao X et al. Computational Platform Based on Deep Learning for Segmenting Ventricular Endocardium in Long-axis Cardiac MR Imaging. Annu Int Conf IEEE Eng Med Biol Soc 2018;2018:4500-4503.

27. Muraru D, Badano LP, Nagata Y et al. Development and prognostic validation of partition values to grade right ventricular dysfunction severity using 3D echocardiography. *Eur Heart J Cardiovasc Imaging* 2020;21:10-21.
28. Shad R, Quach N, Fong R et al. Predicting post-operative right ventricular failure using video-based deep learning. *Nature Communications* 2021;12:5192.

Figure Legends

Figure 1 Flowchart of patient and video selection

3D – three-dimensional, A4C – apical 4-chamber, FPS – frame per second

Figure 2 Correlation between the ground truth and DL-predicted RVEF values in the internal and the external validation sets

Scatterplots of the upper panels show per-video results, whereas those of the lower panels depict per-examination results.

DL – deep learning, RVEF – right ventricular ejection fraction

Figure 3 Receiver operating characteristic curves illustrating the performance of deep learning-predicted right ventricular ejection fraction values in identifying right ventricular dysfunction

95% confidence intervals were calculated from 10,000 stratified bootstrap resamples.

AUC – area under the receiver operating characteristic curve

Figure 4 Kaplan-Meier curves illustrating the event-free survival of patient subgroups in the external validation set

Subgroups were created by splitting the external set based on the DL-predicted, the 3D echocardiography-derived (i.e., ground truth), and the CMR imaging-derived RVEF values.

In the upper panels, the standard cut-off value of 45% was used for dividing patients into subgroups, whereas in the lower panels, patients were split into four quartiles based on the RVEF values. Same-day CMR imaging examination was performed in 354 (97.0%) patients.

CMR – cardiac magnetic resonance, DL – deep learning, RVEF – right ventricular ejection fraction

Figure 5 Saliency maps of three representative cases

The first case (Case #1, Supplementary Video 1) is a 39-year-old healthy, sedentary female with an LVEF of 65.9% and an RVEF of 62.8% measured by 3D echocardiography. Our tool

predicted an RVEF of 62.7% using her apical 4-chamber video. Her heatmaps showed that the dynamics of the atrial and ventricular myocardium, mitral and tricuspid annuli, and valves provided important information for the system, while extracardiac structures did not. The second case (Case #2, Supplementary Video 2) is a 56-year-old male with end-stage heart failure and a cardiac resynchronization therapy defibrillator device. LVEF was 10.8%, RVEF was 27.5%, and the predicted RVEF was 30.7%. Notably, the tricuspid annular plane systolic excursion was only slightly below the reference range (16 mm). On his heatmap, dominantly, the mitral and tricuspid annuli, the myocardium of the RV, and the right atrium provided relevant information for the system in systole. The third case (Case #3, Supplementary Video 3) is another heart failure patient (84-year-old male) with an LVEF of 22.4% but an RVEF of 48.3%. Our tool predicted an RVEF of 45.3%. On his heatmap, principally, the interventricular septum and the tricuspid annulus provided important information in systole, while the LV myocardium confused the system.

DL – deep learning, LV – left ventricular, LVEF – left ventricular ejection fraction, RV – right ventricular ejection fraction

Central Illustration A fully automated deep learning-based tool for the prediction of 3D echocardiography-derived right ventricular ejection fraction

Using a single-center database of patients with 2D A4C view echocardiographic videos and 3D echocardiography-derived RVEF values, we implemented a deep learning-based ensemble model to predict RVEF values from 2D videos alone. We validated the proposed model both internally and externally. In the external dataset, we also benchmarked the model's performance against human expert reading and explored the association of the predicted RVEF values with clinical outcomes.

A4C – apical 4-chamber, RV – right ventricular, RVEF – right ventricular ejection fraction

Table 1 Clinical, echocardiographic, and cardiac magnetic resonance imaging characteristics of the external validation cohort

	All n=365	MACE n=49	No MACE n=316	p-value
Age, years	68 (58-75)	74 (65-79)	67 (57-75)	0.001
Males	237 (64.9)	28 (57.1)	209 (66.1)	0.260
BSA, m ²	1.6 (1.5-1.8)	1.6 (1.5-1.7)	1.6 (1.5-1.8)	0.054
BMI, kg/m ²	22.2 (20.3-25.0)	21.7 (19.3-24.5)	22.2 (20.4-25.0)	0.292
HR, 1/min	67 (59-76)	69 (60-82)	66 (59-74)	0.197
SBP, mmHg	129 (113-146)	116 (107-133)	129 (114-146)	0.003
DBP, mmHg	71 (63-79)	66 (58-71)	72 (64-80)	<0.001
NYHA functional class				
I	143 (39.2)	8 (16.3)	135 (42.7)	
II	183 (50.1)	26 (53.1)	157 (49.7)	
III	38 (10.4)	14 (28.6)	24 (7.6)	
IV	1 (0.3)	1 (2.0)	0 (0.0)	<0.001
Primary indication for CMR imaging				
Ischemic heart disease	123 (33.7)	16 (32.7)	107 (33.9)	
Heart failure	70 (19.2)	12 (24.5)	58 (18.4)	
Primary cardiomyopathy	50 (13.7)	4 (8.2)	46 (14.6)	
Valvular heart disease	38 (10.4)	6 (12.2)	32 (10.1)	
Pulmonary hypertension	10 (2.7)	3 (6.1)	7 (2.2)	
Other	74 (20.3)	8 (16.3)	66 (20.9)	0.371
Risk factors				
Hypertension	202 (55.3)	29 (59.2)	173 (54.7)	0.644
Diabetes	104 (28.5)	18 (36.7)	86 (27.2)	0.177
Hyperlipidemia	159 (43.6)	23 (46.9)	136 (43.0)	0.644
Coronary artery disease	150 (41.1)	19 (38.8)	131 (41.5)	0.757
Chronic kidney disease	155 (42.5)	30 (61.2)	125 (39.6)	0.005
Echocardiography				
3DE LVEDV, mL	143 (112-197)	175 (136-206)	138 (109-194)	0.005
3DE LVESV, mL	82 (56-134)	119 (80-153)	79 (53-127)	<0.001
3DE LVEF, %	41 (30-51)	31 (24-43)	43 (32-51)	<0.001
3DE LVGLS, %	-12 ((-8)-(-16))	-9 ((-6)-(-12))	-13 ((-9)-(-16))	<0.001
Tricuspid s', cm/s	10.4 (8.5-12.2)	9.6 (8.5-11.2)	10.6 (8.5-12.3)	0.154
TAPSE, mm	17 (13-21)	15 (12-19)	17 (13-21)	0.018
Fractional area change, %	37 (30-42)	35 (27-41)	37 (31-43)	0.029
2DE RVFWLS, %	-19 ((-23)-(-15))	-16 ((-19)-(-11))	-20 ((-24)-(-16))	<0.001
3DE RVEDV, mL	100 (82-121)	113 (88-135)	98 (80-117)	0.010
3DE RVESV, mL	53 (39-67)	63 (53-81)	51 (38-65)	<0.001
3DE RVEF, %	48 (40-54)	41 (34-49)	49 (42-55)	<0.001
CMR imaging				
CMR LVEDV, mL*	157 (118-210)	186 (150-234)	150 (117-204)	0.005
CMR LVESV, mL*	94 (61-144)	137 (83-195)	86 (58-138)	<0.001
CMR LVEF, %*	39 (26-51)	27 (17-40)	41 (27-52)	<0.001

CMR RVEDV, mL*	105 (88-136)	108 (88-141)	105 (88-134)	0.752
CMR RVESV, mL*	58 (45-80)	68 (52-84)	57 (44-79)	0.090
CMR RVEF, %*	44 (37-51)	39 (31-47)	45 (38-52)	0.001

*Same-day CMR imaging examination was performed in 354 (97.0%) patients.

Continuous variables are presented as median (interquartile range) and categorical variables as n (%). Comparisons between patients with and without MACE were performed using unpaired Student's t-test or Mann-Whitney U test for continuous variables, Chi-squared or Fisher's exact test for categorical variables, as appropriate.

2DE – two-dimensional echocardiography, 3DE – three-dimensional echocardiography, BMI – body mass index, BSA – body surface area, CMR – cardiac magnetic resonance, DBP – diastolic blood pressure, HR – heart rate, LVEDV – left ventricular end-diastolic volume, LVEF – left ventricular ejection fraction, LVESV – left ventricular end-systolic volume, LVGLS – left ventricular global longitudinal strain, MACE – major adverse cardiac event, NYHA – New York Heart Association, RVEDV – right ventricular end-diastolic volume, RVESV – right ventricular end-systolic volume, RVFWLS – right ventricular free-wall longitudinal strain, SBP – systolic blood pressure, TAPSE – tricuspid annular plane systolic excursion

Table 2 Independent predictors of major adverse cardiac events identified using univariable and multivariable Cox regression

	Univariable		Multivariable					
	HR [95% CI]	p-value	Model 1		Model 2		Model 3	
	HR [95% CI]	p-value	HR [95% CI]	p-value	HR [95% CI]	p-value	HR [95% CI]	p-value
Age	1.054 [1.025-1.084]	<0.001	1.062 [1.031-1.094]	<0.001	1.058 [1.028-1.089]	<0.001	1.066 [1.035-1.099]	<0.001
Male sex	0.764 [0.433-1.350]	0.354	0.653 [0.363-1.175]	0.155	0.605 [0.332-1.103]	0.101	0.627 [0.345-1.138]	0.125
3DE LVEF	0.960 [0.940-0.980]	<0.001	0.979 [0.947-1.012]	0.210	0.972 [0.946-0.998]	0.037		
CMR LVEF*							0.971 [0.950-0.993]	0.009
DL-predicted RVEF	0.912 [0.875-0.950]	<0.001	0.924 [0.862-0.990]	0.025				
Ground truth RVEF	0.936 [0.911-0.962]	<0.001			0.951 [0.917-0.986]	0.007		
CMR RVEF*	0.966 [0.943-0.989]	0.005					0.967 [0.938-0.998]	0.036

Same-day CMR imaging examination was performed in 354 (97.0%) patients.

3DE – three-dimensional echocardiography, CI – confidence interval, CMR – cardiac magnetic resonance, DL – deep learning, HR – hazard ratio, LVEF – left ventricular ejection fraction, RVEF – right ventricular ejection fraction

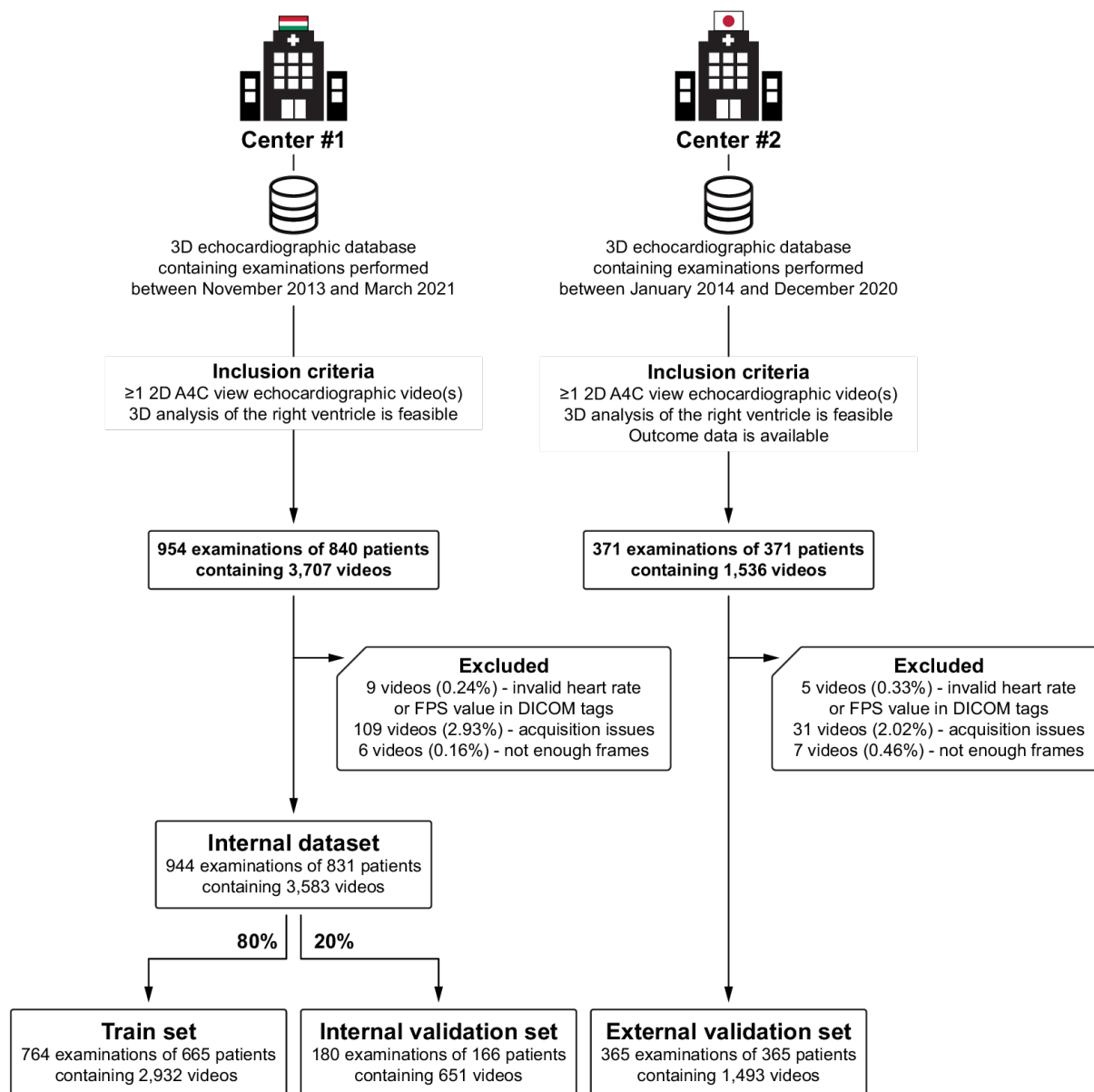


Figure 1: Flowchart of patient and video selection

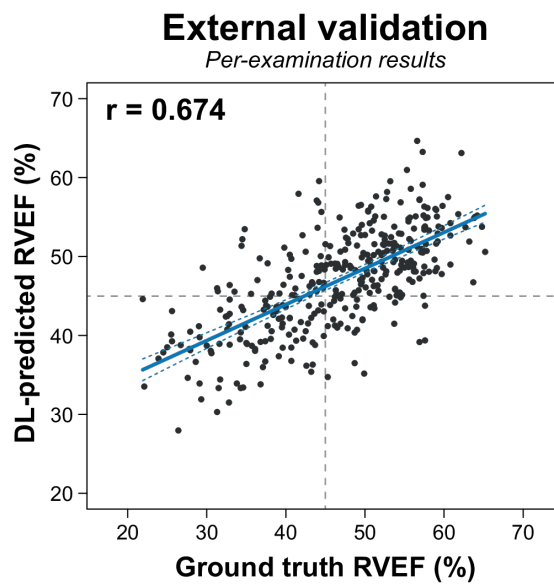
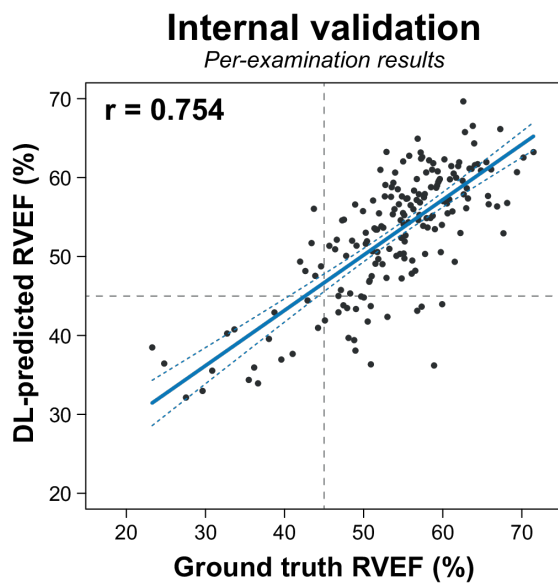
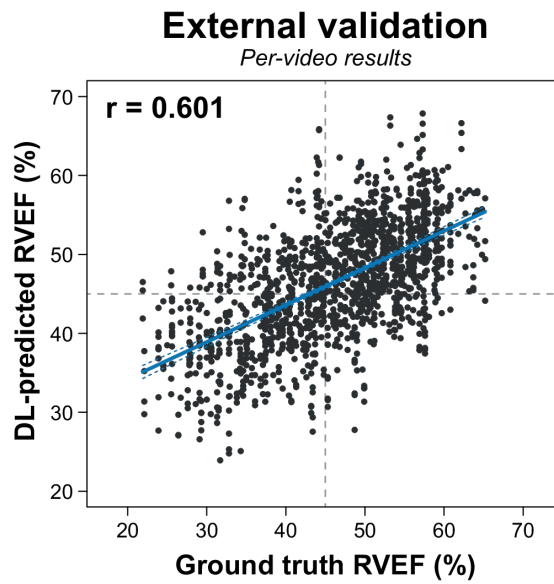
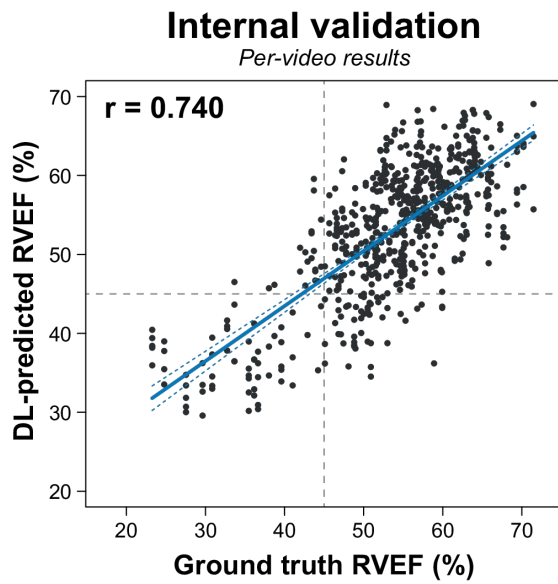


Figure 2: Correlation between the ground truth and DL-predicted RVEF values in the internal and the external validation sets

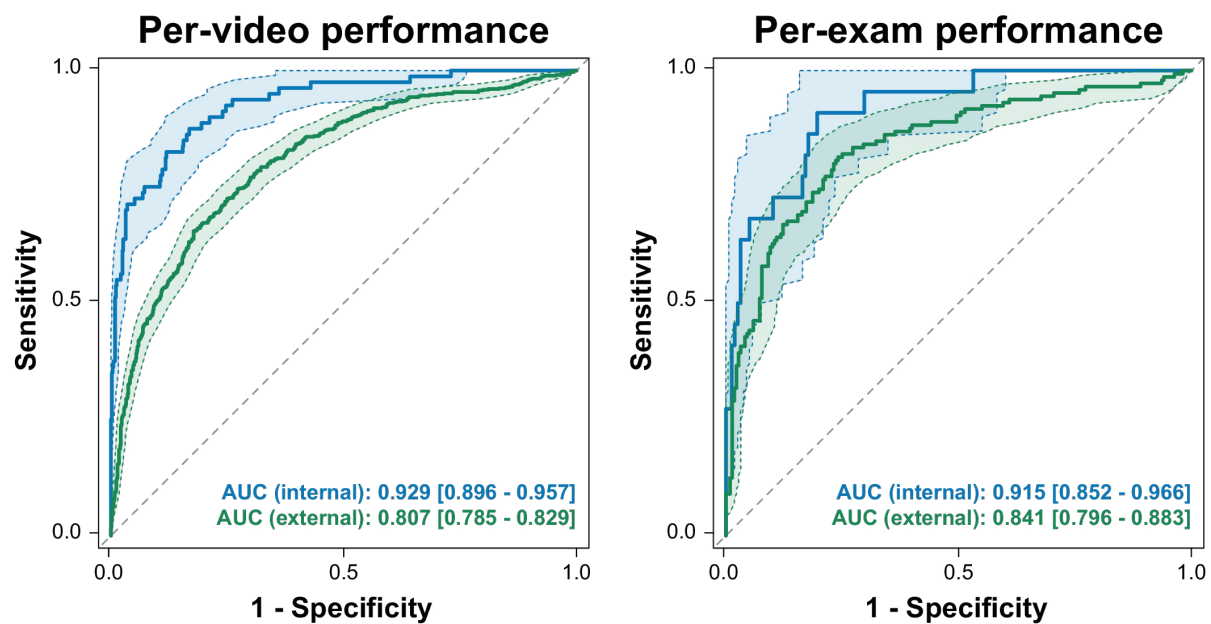


Figure 3: Receiver operating characteristic curves illustrating the performance of deep learning-predicted right ventricular ejection fraction values in identifying right ventricular dysfunction

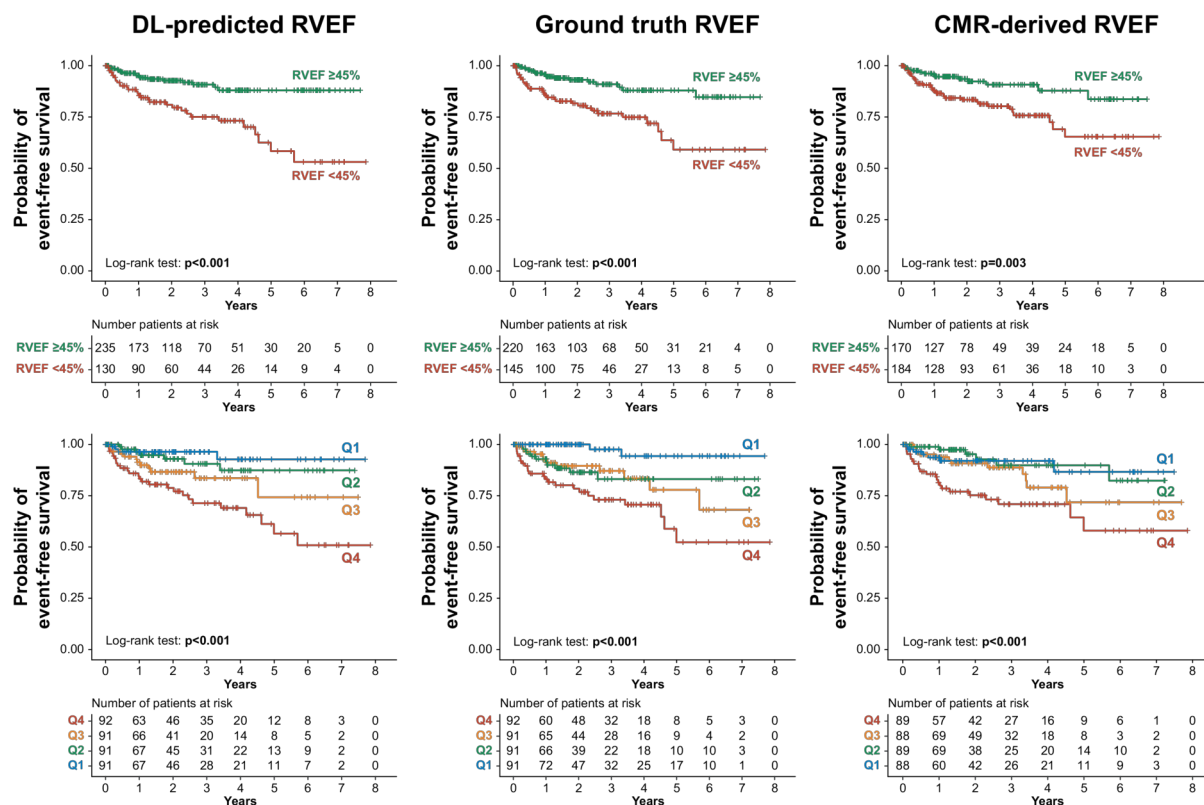


Figure 4: Kaplan-Meier curves illustrating the event-free survival of patient subgroups in the external validation set

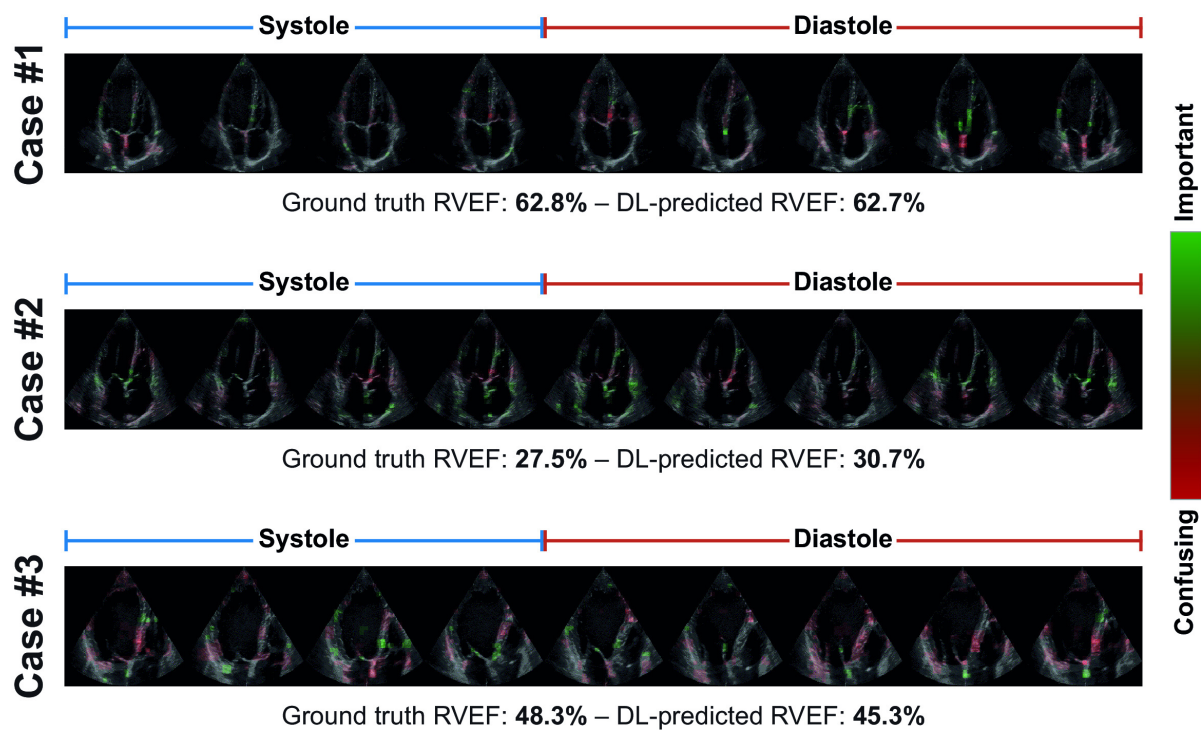


Figure 5: Saliency maps of three representative cases

