RVENet: A Large Echocardiographic Dataset for the Deep Learning-Based Assessment of Right Ventricular Function

Bálint Magyar*[1], Márton Tokodi*[2,3], András Soos[1,2], Máté Tolvaj[2], Bálint Károly Lakatos[2], Alexandra Fábián[2], Elena Surkova[4], Béla Merkely#[2], Attila Kovács#[2], András Horváth#[1]

[1] Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Budapest, Hungary magyar.balint@itk.ppke.hu
[2] Heart and Vascular Center, Semmelweis University, Budapest, Hungary attila.kovacs@med.semmelweis-univ.hu
[3] Division of Cardiovascular Diseases and Hypertension, Rutgers Robert Wood Johnson Medical School, New Brunswick NJ 08901, USA
[4] Harefield Hospital, Royal Brompton and Harefield Hospitals, Part of Guy's and St Thomas' NHS Foundation Trust, London, United Kingdom

*These authors contributed equally to this work and are joint first authors.
#These authors contributed equally to this work and are joint last authors.

Abstract. Right ventricular ejection fraction (RVEF) is an important indicator of cardiac function and has a well-established prognostic value. In scenarios where imaging modalities capable of directly assessing RVEF are unavailable, deep learning (DL) might be used to infer RVEF from alternative modalities, such as two-dimensional echocardiography. For the implementation of such solutions, publicly available, dedicated datasets are pivotal.
Accordingly, we introduce the RVENet dataset comprising 3,583 two-dimensional apical four-chamber view echocardiographic videos of 831 patients. The ground truth RVEF values were calculated by medical experts using three-dimensional echocardiography. We also implemented benchmark DL models for two tasks: (i) the classification of RVEF as normal or reduced and (ii) the prediction of the exact RVEF values. In the classification task, the DL models were able to surpass the medical experts' performance. We hope that the publication of this dataset may foster innovations targeting the accurate diagnosis of RV dysfunction.

Keywords: Echocardiography, Right ventricle, Right ventricular ejection fraction, Deep learning

1 Introduction

Echocardiography is an ultrasound-based imaging modality that aims to study the physiology and pathophysiology of the heart. Important indicators that describe the cardiac pump function can be calculated based on the annotations

of echocardiographic recordings. One of these indicators is the EF. This ratio indicates the amount of blood pumped out by the examined ventricle during its contraction. In other words, EF defines the normalized difference between the end-diastolic volume (EDV) which is the largest volume during the cardiac cycle, and the end-systolic volume (ESV) which is the smallest volume during the cardiac cycle.

$$EF(\%) = \frac{EDV - ESV}{EDV} * 100$$

Two-dimensional (2D) echocardiographic images acquired from standardized echocardiographic views can be used to approximate left ventricular volumes, and thus, the left ventricular ejection fraction (LVEF) with sufficient accuracy. Due to the more complex three-dimensional (3D) shape of the RV (see Figure 1), there is no accurate and clinically used method for estimating RV ejection fraction (RVEF) from 2D recordings [15]. Nevertheless, the availability of this indicator in daily clinical practice would be highly desirable. RVEF can be calculated using 3D echocardiography, which is validated against the gold-standard cardiac magnetic resonance imaging; however, it requires additional hardware and software resources along with significant human expertise to maximize its accuracy and reproducibility [11].

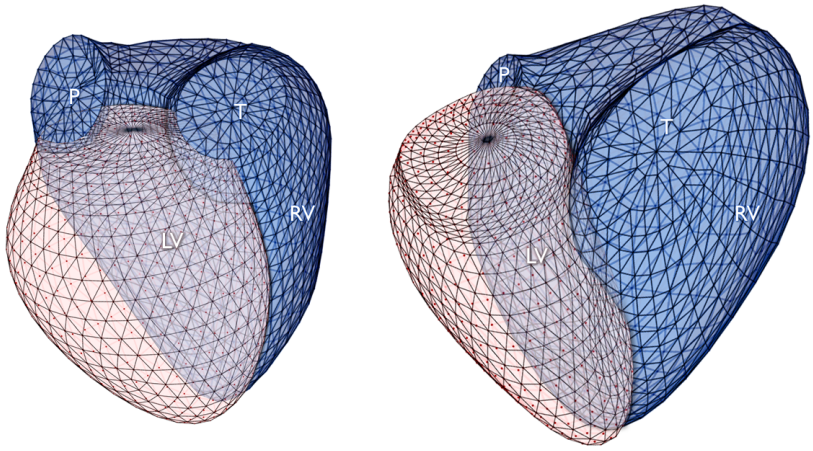

Fig. 1: This figure depicts the left and right ventricles of a normal (left) and a diseased heart (right), as well as the geometric differences between the two ventricles. It can be seen that the estimation of the left ventricular volume from a single 2D plane is feasible due to its regular shape. In contrast, the right ventricle's more complex shape requires 3D evaluation. RV - right ventricle, LV - left ventricle, P - pulmonary valve, T - tricuspid valve

In this work, we publish a dataset that contains 2D echocardiographic videos from 831 patients suitable for RV function assessment. 3D recordings were also collected from the same patients for ground truth generation (i.e., RVEF). We assume that deep learning methods can find relevant patterns on the 2D recordings to detect RV dysfunction and to predict the exact value of RVEF.

To test our assumption, we used an off-the-shelf video classification model and a custom spatiotemporal convolutional neural network for the classification of reduced/normal RVEF and for the prediction of RVEF. RV function can be classified as normal if the RVEF is equal or above 45%, and reduced if it is below 45%. Even if the prediction of RVEF seems to be a more comprehensive method, the aforementioned binary classification is the most widely used evaluation in clinical practice having a clear association with the risk of future adverse clinical events [6, 10].

Our contributions can be summarized as follows:

- We publish a large-scale echocardiographic dataset for the assessment of RV function. According to our knowledge, this is the first dedicated dataset aiming RV evaluation. Its uniqueness lies in the calculation of the ground truth RVEF which was done using 3D recordings.
- Baseline deep learning models were developed and applied to classify RV reduced/normal EF and to predict the RVEF value. Based on our literature review, there is no solution that solves the same clinical problem.
- We compared the models' performance with two experienced medical doctors' performance (one from the center from which the dataset originates and one external expert). This was a unique comparison, as the ground truth values were created using another modality (3D echocardiography).

2 Related Work

2.1 Datasets

To the best of our knowledge, there is no dataset for RV function assessment. Based on an exhaustive literature review the following open-source datasets for the assessment of LV function were identified.

The first one called CETUS [3] which contains 45 3D echocardiogram recordings (each from a different patient). The dataset was collected from three different hospitals using different machines, and it was annotated (3D LV segmentation) by three expert cardiologists based on a pre-defined protocol. The main purpose of this dataset is to compare 3D segmentation algorithms, therefore different segmentation metrics (e.g. Dice similarity index and 3D Hausdorff distance) are used for evaluation. To evaluate the clinical performance of the methods they calculate mean absolute error (MAE) and means squared error (MSE) for the ESV, EDV and LVEF values.

CAMUS represents another important dataset [7]. 500 2D echocardiogram recordings (each from a different patient) were collected from a single hospital for this data set. LV segmentation mask that was created by an experienced

cardiologist (except in the test set where 2 other cardiologist were involved) and LVEF values are provided for the recordings. The estimation of the LVEF values based on the Simpson's biplane method of discs [5]. The same measures were applied for evaluation as in the CETUS dataset (except the 2D version of segmentation metrics instead of 3D).

The third dataset is the EchoNet-Dynamic [13] which contains 10036 2D echocardiogram recordings (each from a different patient). Each recording was captured from apical-4-chamber view. LVEF, EDV and ESV measurements were obtained by experienced cardiologists on the standard clinical workflow. The original recordings were filtered based on image quality and resized to 112x112 gray-scale image sequences.

They also used MAE and MSE as well as R^2 to evaluate the LVEF, EDV and ESV predictions.

2.2 Methods

The assessment of cardiac functions using machine learning and particularly deep learning algorithms is a commonly applied approach nowadays due to the superior performance of these systems compared to traditional computer vision algorithms. These methods can even surpass human performance in certain cases [2][22][1].

Some of the earlier methods focused on individual frames to predict echocardiographic view, ventricular hypertrophy, EF, and other metrics. Madani et. al implemented a method for view classification, which is usually the first step before further analysis [9].

Another group of researchers developed a more complex system that identifies the view, applies segmentation on the 2D recordings and predicts the LVEF as well as one of 3 diagnostic classes [23].

Leclerc et. al used a more advanced encoder-decoder style segmentation network on the CAMUS dataset to calculate the EDV and ESV and to predict the LVEF using these values [7].

Application of anatomical atlas information as a segmentation constraint was successfully applied by Oktay et al. to create a more accurate segmentation method for LV segmentation and LVEF prediction [12].

In order to predict measures like EF that can be estimated only using multiple frames, the key frames has to be selected manually for these methods.

More advanced methods use video input and spatiotemporal convolutional networks to provide an end-to-end solution.

Shat et. al applied 3D convolutional layers along with optical flow to detect temporal changes along the video and to predict post-operative RV failure [18].

The effect of temporally consistent segmentation has been investigated previously [4]. In this study, the authors used a custom convolutional layer to obtain bi-directional motion fields. The motion detection was combined with the segmentation results to obtain precise LV segmentation.

Ouyang et al. [14] presented a two stage convnet applying atrous convolution to first segment the LV, and then another stage of spatiotemporal convolutions to predict the LVEF.

Compared to existing approaches, we proposed a single stage method that aims to predict the RVEF directly from the input videos using satiotemporal convolutional networks. We assume that in contrast to segmentation based methods, different regions of the input image can also contribute to the RVEF prediction, and therefore this task is feasible.

3 Overview of the RVENet Dataset

3.1 Data Collection

To create the RVENet dataset, we retrospectively reviewed the transthoracic echocardiographic examinations performed between November 2013 and March 2021 at the Heart and Vascular Center of Semmelweis University (Budapest, Hungary). We aimed to identify those examinations that included one or more 2D apical four-chamber view echocardiographic videos and an electrocardiogram (ECG)-gated full-volume 3D echocardiographic recording (with a minimum volume rate of 15 volumes/second, acquired from an RV-optimized apical view, and reconstructed from four cardiac cycles) suitable for 3D RV analysis and RVEF assessment. The 2D apical four-chamber view videos were exported as Digital Imaging and Communications in Medicine (DICOM) files, whereas the 3D recordings were used for generating labels (see the detailed description of data labeling in section 3.2). Protected health information was removed from all exported DICOM files. 2D videos with (i) invalid heart rate or frame per second (FPS) values in the DICOM tags, (ii) acquisition issues comprising but not limited to severe translational motion, gain changes, depth changes, view changes, sector position changes, (iii) duration shorter than one cardiac cycle, or (iv) less than 20 frames per cardiac cycle were discarded. All transthoracic echocardiographic examinations were performed 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).

3.2 Data Labeling

The exported 2D echocardiographic videos were reviewed by a single experienced echocardiographer who (i) assessed the image quality using a 5-point Likert scale (1 - non-diagnostic, 2 - poor, 3 - moderate, 4 - good, 5 - excellent), (ii) labeled them as either standard or RV-focused, (iii) determined LV/RV orientation (Mayo - RV on the right side and LV on the left side; Stanford - LV on the right side and RV on the left side), and (iv) ascertained that none of them meet the exclusion criteria.

The 3D echocardiographic recordings were analyzed by expert readers on desktop computers using a commercially available software solution (4D RV-Function 2, TomTec Imaging, Unterschleissheim, Germany) to compute RV end-diastolic and end-systolic volumes, as well as RVEF. These parameters were calculated only once for each echocardiographic examination. However, an examination may contain multiple 2D apical four-chamber view videos; thus, the same label was linked to all 2D videos within that given examination.

A comprehensive list and description of the generated labels are provided in Table 1.

Variable	Description
FileName	Hashed file name used to link videos and labels
PatientHash	Hashed patient name
PatientGroup	Patient subgroup referring to the primary diagnosis
Age	Age in years, rounded to nearest year
Sex	Sex reported in medical record (M - male, F - female)
UltrasoundSystem	Ultrasound system used for video acquisition
FPS	Frames per second (1/s)
NumFrames	Number of frames in the whole video
VideoViewType	Standard or RV-focused apical four-chamber view
VideoOrientation	LV/RV orientation (Mayo or Stanford)
VideoQuality	2D video quality on a 5-point scale (1 - non-diagnostic, 2 - poor, 3 - moderate, 4 - good, 5 - excellent)
RVEDV	3D echocardiography-derived RV end-diastolic volume (mL)
RVESV	3D echocardiography-derived RV end-systolic volume (mL)
RVEF	3D echocardiography-derived RV ejection fraction (%)
Split	Train-test splitting used for benchmarking

Table 1: Description of the labels. RV - right ventricular

3.3 Composition of the Dataset

The RVENet dataset contains 3,583 2D apical four-chamber view echocardiographic videos from 944 transthoracic echocardiographic examinations of 831 individuals. It comprises ten distinct subgroups of subjects: (i) healthy adult volunteers (without history and risk factors of cardiovascular diseases, n=192), (ii) healthy pediatric volunteers (n=54), (iii) elite, competitive athletes (n=139), (iv) patients with heart failure and reduced EF (n=98), (v) patients with LV non-compaction cardiomyopathy (n=27), (vi) patients with aortic valve disease (n=85), (vii) patients with mitral valve disease (n=70), (viii) patients who underwent orthotopic heart transplantation (n=87), (ix) pediatric patients who underwent kidney transplantation (n=23), and (x) others (n=56). Beyond the primary diagnosis and the labels mentioned in section 3.2, we also provided the

age (rounded to the nearest year) and the biological sex (as reported in medical records) for each patient, and train-test splitting (80:20 ratio) that we used for the training and the evaluation of the benchmark models (see section 4). In addition, the ultrasound system utilized for video acquisition, the frame rate (i.e., FPS), and the total number of frames are also reported for each video among the labels.

3.4 Data De-identification and Access

Before publication of the RVENet dataset, all DICOM files were processed to remove any protected health information. We also ensured that no protected health information is included among the published labels. Thus, the RVENet dataset complies with the General Data Protection Regulation of the European Union. The dataset with the corresponding labels is available at https://rvenet.github.io/dataset/.

The RVENet dataset is available only for personal, non-commercial research purposes. Any commercial use, sale, or other monetization is prohibited. Re-identification of individuals is strictly prohibited. The RVENet dataset can be used only for legal purposes.

4 Benchmark Models

4.1 Methodology

Ethical Approval The study conforms to 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). Methods and results are reported in compliance with the Proposed Requirements for Cardiovascular Imaging-Related Machine Learning Evaluation (PRIME) checklist (Supplementary Table 3) [17].

Data Preprocessing The RVENet dataset can be used for various purposes within the realm of cardiovascular research. In this section, we describe data preprocessing that proceeded both deep learning tasks, namely (i) the prediction of the exact RVEF values (i.e. regression task) and (ii) and the classification of reduced/normal RVEF (i.e. binary classification task).

All echocardiographic recordings were exported as DICOM files. Each DICOM file contains a series of frames depicting one or more cardiac cycles. This arrangement of the data had to be modified to achieve a representation that is more suitable for neural networks. The three main steps of preprocessing can be described as follows: (1) frame selection, (2) image data preparation, and (3) handling imbalance in the train set.

Frame selection refers to the preprocessing step in which 20 frames are selected to represent a cardiac cycle (20 frames per cardiac cycle proved to be the appropriate number in [14] for left ventricle EF prediction). Recordings may

contain multiple cardiac cycles and may differ in length and frame rate (FPS - frames per second). We applied the following formula to estimate the length of a cardiac cycle (L) based on heart rate (HR) and FPS extracted from the DICOM file tags:

$$L = \frac{60}{HR} * FPS$$

Since all the recordings are ECG-gated and start with the end-diastolic frame, the split of the videos into consecutive, non-overlapping fragments (depicting exactly one cardiac cycle) was feasible. Fragments containing less than L frames were excluded. Then, a predefined number of frames (N = 20) were sampled from the fragments based on the sampling frequency (SF) which was calculated using the following formula:

$$SF = \frac{L}{N}$$

A subset of randomly selected videos underwent a manual verification process by an experienced physician to evaluate the cardiac cycle selection.

The next step is the image data generation which is shown in Figure 2. The selected frames contain multiple components that are unnecessary for the neural network training, such as ECG signal, color-scale and other signals and texts. These unwanted items were removed using motion-based filtering. Our algorithm tracks the changes frame by frame and set the pixels to black if they change fewer times than a predefined threshold (in our case 10). We also cropped the relevant region of the recordings and generate a binary mask for training.

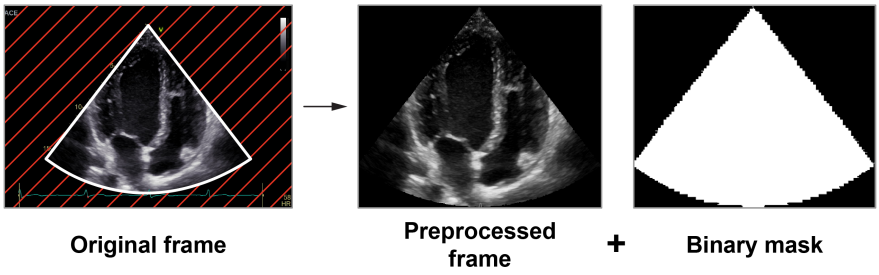

Fig. 2: Schematic illustration of the preprocessing. First, static objects (e.g., technical markers and the ECG signal) were removed from the area marked by red diagonal lines using motion-based filtering, while the region of interest (enclosed by the white contour) was left intact. Second, the region of interest was cropped from the filtered image and a binary mask was also generated.

The removal of unwanted components is performed along with the binary mask generation. This mask is created for every video fragment with the consideration of all the frames. The aim of this additional binary image is to prevent the network from extracting features from the outside of the region of interest.

The closest enclosing rectangle is applied to both the preprocessed frames and the binary image. After that they are resized to 224x224 pixels. Previously, Madani et al. examined the effect of the input image size on the system's accuracy in a echocardiography view classification task , and they found out that the accuracy saturated if they used higher resolution than this [9].

Train-Test Splitting The dataset was split in an approximately 80:20 ratio into train and test 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 test set).

Dataset Balancing An optional step in pre-processing is the dataset balancing which aims to compensate the high imbalance between negative (normal EF) and positive (reduced EF) cases. In binary classification this means the oversampling of the positive cases and the undersampling of negative cases. In case of a regression problem, the EF values are assigned to discrete bins, and the algorithm aims to balance the number of samples in these bins. The method takes the number of videos and heart cycles from a certain patient into account. It aims to keep at least one video from every patient in the undersampling phase, and oversamples the videos from patients with reduced EF uniformly.

Spatiotemporal Convolutional Neural Networks As it was mentioned in the Related Work section, spatiotemporal processing of the echocardiographic videos provide a more accurate approach for EF prediction [4, 14].

Based on these results, we used two neural network models. The first one is composed of R(2+1)D spatiotemporal convolutional blocks [20], and a PyTorch implementation (called R2Plus1D_18) of such a model is available off-the-shelf. We refer to this model as "R2+1D" in the text. We also designed a more efficient, single stage neural network called EFNet (Ejection Fraction Network) that consist of a feature extractor backbone (ShuffleNet V2 [8] or ResNext50 [21]) a temporal merging layer and two fully connected layers. The architecture of our custom model is visualized in (Figure 3).

Both networks predict the RVEF directly from an input image sequence (and the corresponding binary mask). The same architectures can be used for classification or regression by changing the number of outputs.

Model training and evaluation Several experiments were performed to find the best training parameters. In the followings, the final parameter sets are introduced.

The backbone model of the EFNet was ShuffleNet V2 [8] for binary classification and ResNext50 [23] for regression, which is a more challenging task and therefore needs a more complex architecture. To distinguish these two versions of EFNet, we refer to the classification model as EFNet_class and to the regression model as EFNet_reg.

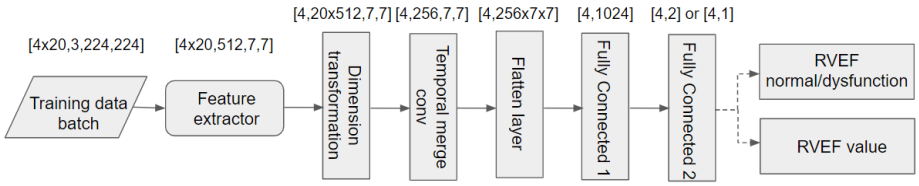

Fig. 3: The training batch contains batch size × video frames images with resolution of 224x224 pixels. The feature extractor is a ShuffleNet V2 [8] or a ResNext50 [23] model. The dimension transformation layer groups the frame features corresponding to the videos in the batch, then these group of frames are processed using the spatiotemporal convolutional layer to extract dynamic features. The final features are downscaled using fully connected layers and forwarded to either a classification or a regression head.

ImageNet pre-trained weights can improve the performance of deep learning models applied in medical datasets [19]. In our case, only the EFNet_reg model was initialized using ImageNet pre-trained weights, the weights of the R2+1D and EFNet_reg models were initialized randomly.

As it was described in 4.1, the videos were split into distinct cardiac cycles, and 20 frames were sampled from each of them.

Dataset balancing (described above) also improved the accuracy of the system as well as the F1 and R^2 scores. This is mainly due the substantial imbalance in the dataset.

Augmentation techniques were also applied, namely vertical flipping and rotation (+/- 10°). Normalization was not applied.

The models were trained for maximum 30 epochs with a batch size of 4. We used Adam optimizer (initial learning rate = 0.003, momentum = 0.9), and the PyTorch cyclic learning rate scheduler (lambda=0.965).

Cross-entropy loss was used in the classification experiments, and MAE in the regression experiments.

The parameter search and model selection was done applying a four fold cross validation using the whole training set. For the final experiments, the training set was split in 75%-25% ratio into training and validation set. A balanced version of the training set was used for training, and the best model was selected using the validation set results based on f1 score in case of classification and R^2 in case of regression training.

For both the classification and regression tasks, the deep learning models were evaluated on the test set. As a classification model predicts a class for a single cardiac cycle, these predictions were averaged for each video (taking the majority vote). This way the results can be compared with the human experts' performance as they also saw the whole video during evaluation. In the regression task, the models' prediction were averaged for each test video similarly to the classification task. In this case no human expert comparison was performed as

the exact prediction of the RVEF value is not part of the clinical evaluation of 2D echocardiographic recordings (it is done only using 3D recordings).

Human Expert Evaluation Videos of the test set were evaluated by two expert cardiologists: one from the same center where the echocardiographic videos were acquired (referred to as $\text{Expert}_{\text{Internal}}$) and one from an external center (referred to as $\text{Expert}_{\text{External}}$). Although patient identification information, medical history, and diagnosis were hidden from both of them during evaluation, the first cardiologist might have seen some of the videos previously, as he performs echocardiographic examinations at a daily basis. On the other hand, the second cardiologist has not seen any of the videos previously, enabling a completely unbiased comparison.

Both evaluator used the same custom desktop application for evaluation, which displayed the original videos one by one in a random order and the evaluating expert had to decide based on visual estimation whether the video belongs to patient with normal (RVEF is equal or greater than 45) or reduced (RVEF is less than 45) RV function.

4.2 Results

Table 2 shows the results of the deep learning models and the human experts in the detection of RV dysfunction (i.e., binary classification task). Both deep learning models achieved a numerically higher accuracy, specificity, sensitivity, and F1 score than the medical experts. We also confirmed these differences using McNemar's tests. EFNet_class model exhibited an accuracy, specificity, and sensitivity comparable to those of the internal expert, whereas it had higher accuracy and sensitivity than the external expert (Table 3). The R2+1D model achieved a higher sensitivity than the internal expert, and it also outperformed the external expert in terms of accuracy and sensitivity (Table 3).

	Accuracy	**Specificity**	**Sensitivity**	**F1 score**
EFNet_class	0.911	0.942	0.688	0.655
R2+1D	0.920	0.940	0.775	0.705
$\text{Expert}_{\text{Internal}}$	0.897	0.940	0.588	0.584
$\text{Expert}_{\text{External}}$	0.859	0.923	0.400	0.410

Table 2: Performance of the deep learning models and the cardiologists in the binary classification task.

In the regression task (i.e. prediction of the exact RVEF value), the two deep learning models performed similarly (Supplementary Figure 1). The EFNet_reg model predicted RVEF with an R^2 of 0.411, a mean absolute error of 5.442 percentage points, and a mean squared error of 47.845 percentage points^2, whereas

	Accuracy	Specificity	Sensitivity
EFNet_class vs. $\text{Expert}_{\text{Internal}}$	0.362	1.000	0.170
EFNet_class vs. $\text{Expert}_{\text{External}}$	0.001	0.229	<0.001
R2+1D vs. $\text{Expert}_{\text{Internal}}$	0.106	1.000	0.004
R2+1D vs. $\text{Expert}_{\text{External}}$	<0.001	0.275	<0.001

Table 3: P-values of the McNemar's tests comparing the accuracy, specificity, and sensitivity of the deep models with those of the cardiologists in the binary classification task.

the R2+1D model achieved an R^2 of 0.417, a mean absolute error of 5.374 percentage points, and a mean squared error of 47.377 percentage points^2.

The Bland-Altman analysis showed a significant bias between the deep learning-predicted and the 3D echocardiography-based ground truth RVEF values (EFNet_reg: -2.496 percentage points, p<0.001; R2+1D: 0.803 percentage points, p<0.001; Supplementary Figure 1).

Table 4 shows the comparison of the R2+1D and the two EFNet models in terms of size, and inference speed. Even if the R2+1D model performed better in the classification task and slightly better in the regression task, EFNet is a more efficient model, and its speed can be a huge advantage in model training and inference both in experimentation and in clinical applications.

	Feature extractor	Inference time [ms]	Model size [MB]
EFNet_class	ShuffleNet V2	16	61
EFNet_reg	ResNext50	31	177
R2+1D	R(2+1)D	53	119

Table 4: Size and inference speed results of the baseline models. Inference speed was measured by averaging 100 iterations with batch size of 1 on an Nvidia V100 GPU.

5 Discussion

5.1 Potential Clinical Application

RV dysfunction is significantly and independently associated with symptomatology and clinical outcomes (e.g. all-cause mortality and/or adverse cardiopulmonary outcomes) in different cardiopulmonary diseases irrespective of which

side of the heart is primarily affected. Among echocardiographic parameters, 3D echocardiography-derived quantification of RVEF provides the highest predictive value for future adverse events [16]. However, there are several issues that prevent RVEF to be a standard measure in the daily clinical routine. 3D echocardiography-based quantification of RVEF requires advanced hardware and software environment along with experienced cardiology specialists. First, a high-end ultrasound system equipped with a 3D-capable matrix transducer is required. Compared to a conventional acquisition of an apical four-chamber view video (included in the routine protocols and takes no more than one minute irrespective of the level of expertise), an RV-focused, modified four-chamber view is needed with the 3D option enabled. The investigator needs to ensure the capture of the entire RV endocardial surface, which can be troublesome with distinct anatomical features of the patient. Moreover, to enable higher temporal resolution, multi-beat reconstruction should be used that can be limited in the cases of irregular heart rhythm, transducer motion, in patients who are not able to breathe-hold, and again, user experience is of significant importance to acquire a high-quality 3D dataset free of artifacts. To acquire such a measurement feasible for RVEF measurement takes about 2 to 4 minutes for an expert user, which can go up to 5-8 minutes for users not having extensive experience in 3D image acquisition. Then, the 3D DICOM file should be post-processed using standalone software (running on a separate PC or embedded in the high-end ultrasound machine). One vendor enables fully automatic 3D reconstruction of the RV endocardial surface and calculation of RVEF values (which takes about 30 seconds). However, in the vast majority of the cases (over 90%), correction of endocardial contours is needed in multiple long- and short-axis views both at the end-diastolic and end-systolic frames. Changes in the automatically traced contours made by the human reader can result in notable interobserver and even intraobserver variability. Here again, the experience of the user is a major factor in terms of accurate measurements and also, analysis time. For an experienced reader, the manual correction of the initial contours takes about 4 to 10 minutes and up to 15 minutes for an inexperienced user. Overall, from image acquisition through image transfer, preprocessing and finally RVEF calculation, the entire process is generally taking 10 to 25 minutes in clinical practice for this single parameter.

Due to significant human and also hardware/software resources needed, RVEF calculation by echocardiography is rarely performed in the clinical routine despite its clear value. This can be circumvented by an automated system, which utilizes routinely acquired echocardiographic videos and does not require a high-end ultrasound system or significant human experience either. In the clinical routine, echocardiography is often performed by other medical disciplines (i.e. emergency physicians, cardiac surgeons) with mobile, even handheld machines to answer focused yet important clinical questions (so-called point-of-care ultrasound examinations). These medical professionals generally do not have any experience with 3D echocardiography and either high-end ultrasound equipment. However, in these disciplines, the detection of RV dysfunction is a critical clin-

ical issue. As it may be applied even to handheld devices and allow the fast detection of RV dysfunction using simple, routine 2D echocardiographic videos, our system could be of high clinical interest. It can run in a cloud environment, and provide results within a few seconds. Also, its use does not require deep technical knowledge.

5.2 Summary

In this paper, we presented a large dataset for the deep learning-based assessment of RV function. We made publicly available 3,583 two-dimensional echocardiographic apical four-chamber view videos from 831 patients to researchers and medical experts. These videos are labelled with RVEF values (the single best echocardiographic parameter for RV function quantification) derived from 3D echocardiography. We also introduced benchmark models, which were able to outperform an external expert human reader in terms of accuracy and sensitivity to detect RV dysfunction. We foresee further performance improvement through collaborations, definition of RV-related specific clinical tasks, addition of further echocardiographic views or imaging modalities. Our current database and model development may serve as a reference point to foster such innovations.

Acknowledgements Project no. RRF-2.3.1-21-2022-00004 (MILAB) has been implemented with the support provided by the European Union.

References

1. Akkus, Z., Aly, Y.H., Attia, I.Z., Lopez-Jimenez, F., Arruda-Olson, A.M., Pellikka, P.A., Pislaru, S.V., Kane, G.C., Friedman, P.A., Oh, J.K.: Artificial intelligence (ai)-empowered echocardiography interpretation: A state-of-the-art review. Journal of Clinical Medicine **10**(7), 1391 (2021)
2. Alsharqi, M., Woodward, W., Mumith, J., Markham, D., Upton, R., Leeson, P.: Artificial intelligence and echocardiography. Echo research and practice **5**(4), R115–R125 (2018)
3. Bernard, O., Bosch, J.G., Heyde, B., Alessandrini, M., Barbosa, D., Camarasu-Pop, S., Cervenansky, F., Valette, S., Mirea, O., Bernier, M., et al.: Standardized evaluation system for left ventricular segmentation algorithms in 3d echocardiography. IEEE transactions on medical imaging **35**(4), 967–977 (2015)
4. Chen, Y., Zhang, X., Haggerty, C.M., Stough, J.V.: Assessing the generalizability of temporally coherent echocardiography video segmentation. In: Medical Imaging 2021: Image Processing. vol. 11596, pp. 463–469. SPIE (2021)
5. Folland, E., Parisi, A., Moynihan, P., Jones, D.R., Feldman, C.L., Tow, D.: Assessment of left ventricular ejection fraction and volumes by real-time, two-dimensional echocardiography. a comparison of cineangiographic and radionuclide techniques. Circulation **60**(4), 760–766 (1979)
6. Lang, R.M., Badano, L.P., Mor-Avi, V., Afilalo, J., Armstrong, A., Ernande, L., Flachskampf, F.A., Foster, E., Goldstein, S.A., Kuznetsova, T., 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. European Heart Journal-Cardiovascular Imaging **16**(3), 233–271 (2015)
7. Leclerc, S., Smistad, E., Pedrosa, J., Østvik, A., Cervenansky, F., Espinosa, F., Espeland, T., Berg, E.A.R., Jodoin, P.M., Grenier, T., et al.: Deep learning for segmentation using an open large-scale dataset in 2d echocardiography. IEEE transactions on medical imaging **38**(9), 2198–2210 (2019)
8. Ma, N., Zhang, X., Zheng, H.T., Sun, J.: Shufflenet v2: Practical guidelines for efficient cnn architecture design. In: Proceedings of the European conference on computer vision (ECCV). pp. 116–131 (2018)
9. Madani, A., Ong, J.R., Tibrewal, A., Mofrad, M.R.: Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease. NPJ digital medicine **1**(1), 1–11 (2018)
10. Muraru, D., Badano, L.P., Nagata, Y., Surkova, E., Nabeshima, Y., Genovese, D., Otsuji, Y., Guida, V., Azzolina, D., Palermo, C., et al.: Development and prognostic validation of partition values to grade right ventricular dysfunction severity using 3d echocardiography. European Heart Journal-Cardiovascular Imaging **21**(1), 10–21 (2020)
11. Muraru, D., Spadotto, V., Cecchetto, A., Romeo, G., Aruta, P., Ermacora, D., Jenei, C., Cucchini, U., Iliceto, S., Badano, L.P.: 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. European Journal of Echocardiography **17**(11), 1279–1289 (2015)
12. Oktay, O., Ferrante, E., Kamnitsas, K., Heinrich, M., Bai, W., Caballero, J., Cook, S.A., De Marvao, A., Dawes, T., O'Regan, D.P., et al.: Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. IEEE transactions on medical imaging **37**(2), 384–395 (2017)

13. Ouyang, D., He, B., Ghorbani, A., Lungren, M.P., Ashley, E.A., Liang, D.H., Zou, J.Y.: Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning. In: NeurIPS ML4H Workshop: Vancouver, BC, Canada (2019)
14. Ouyang, D., He, B., Ghorbani, A., Yuan, N., Ebinger, J., Langlotz, C.P., Heidenreich, P.A., Harrington, R.A., Liang, D.H., Ashley, E.A., et al.: Video-based ai for beat-to-beat assessment of cardiac function. Nature **580**(7802), 252–256 (2020)
15. Porter, T.R., Shillcutt, S.K., Adams, M.S., Desjardins, G., Glas, K.E., Olson, J.J., Troughton, R.W.: Guidelines for the use of echocardiography as a monitor for therapeutic intervention in adults: a report from the american society of echocardiography. Journal of the American Society of Echocardiography **28**(1), 40–56 (2015)
16. Sayour, A.A., Tokodi, M., Celeng, C., Takx, R., Fabian, A., Lakatos, B., Friebel, R., Surkova, E., Merkely, B., Kovacs, A.: Superior prognostic value of three-dimensional echocardiography-derived right ventricular ejection fraction: a meta-analysis. medRxiv (2022)
17. Sengupta, P.P., Shrestha, S., Berthon, B., Messas, E., Donal, E., Tison, G.H., Min, J.K., D'hooge, J., Voigt, J.U., Dudley, J., Verjans, J.W., Shameer, K., Johnson, K., Lovstakken, L., Tabassian, M., Piccirilli, M., Pernot, M., Yanamala, N., Duchateau, N., Kagiyama, N., Bernard, O., Slomka, P., Deo, R., Arnaout, R.: Proposed requirements for cardiovascular imaging-related machine learning evaluation (prime): A checklist. JACC: Cardiovascular Imaging **13**(9), 2017–2035 (2020). https://doi.org/10.1016/j.jcmg.2020.07.015
18. Shad, R., Quach, N., Fong, R., Kasinpila, P., Bowles, C., Castro, M., Guha, A., Suarez, E.E., Jovinge, S., Lee, S., et al.: Predicting post-operative right ventricular failure using video-based deep learning. Nature communications **12**(1), 1–8 (2021)
19. Tajbakhsh, N., Shin, J.Y., Gurudu, S.R., Hurst, R.T., Kendall, C.B., Gotway, M.B., Liang, J.: Convolutional neural networks for medical image analysis: Full training or fine tuning? IEEE transactions on medical imaging **35**(5), 1299–1312 (2016)
20. Tran, D., Wang, H., Torresani, L., Ray, J., LeCun, Y., Paluri, M.: A closer look at spatiotemporal convolutions for action recognition. In: Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. pp. 6450–6459 (2018)
21. Xie, S., Girshick, R., Dollár, P., Tu, Z., He, K.: Aggregated residual transformations for deep neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1492–1500 (2017)
22. Zamzmi, G., Hsu, L.Y., Li, W., Sachdev, V., Antani, S.: Harnessing machine intelligence in automatic echocardiogram analysis: Current status, limitations, and future directions. IEEE reviews in biomedical engineering (2020)
23. Zhang, J., Gajjala, S., Agrawal, P., Tison, G.H., Hallock, L.A., Beussink-Nelson, L., Lassen, M.H., Fan, E., Aras, M.A., Jordan, C., et al.: Fully automated echocardiogram interpretation in clinical practice: feasibility and diagnostic accuracy. Circulation **138**(16), 1623–1635 (2018)