POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

AutoRV: a real-time, fully automated pipeline for RV evaluation based on transesophageal echocardiography in intensive care units

Tesi di Laurea Magistrale in
Biomedical Engineering - Ingegneria Biomedica

Matteo Missana, 247365

Abstract: right ventricle (RV) dysfunction and RV failure are critical complications in intensive care units (ICUs), particularly for patients undergoing mechanical ventilation. While transesophageal echocardiography (TEE) is the preferred imaging modality in these settings, current assessment methods rely on manual, offline analysis, preventing continuous monitoring of RV function, and current automated methods based on deep learning (DL) focus only on transthoracic echocardiography (TTE).

This study introduces AutoRV, a fully automated deep learning pipeline for real-time RV monitoring using 2D TEE mid-esophageal 4-chamber view images. The system was developed to automatically compute clinically relevant indices of RV function based on the automatic detection of three anatomical landmarks: the tricuspid annulus (TA) hinge points (septal and free wall) and the RV apex.

For the detection task, several convolutional neural networks, including ResNet variants, ResNeXt50, SwinUNETR, and U-Net, were trained, tested and validated on a dataset of 8,424 images obtained from 52 mechanically ventilated patients.

The U-Net architecture showed superior performance, with a mean landmark localization error of 4.16 ± 3.36 mm on the test set. Specifically, the detection of the free-wall TA point was highly accurate, with an error of 2.93 ± 1.80 mm. The pipeline proved highly efficient, operating at an inference speed of 248 frames per second (frames per second (*fps*)), making it suitable for real-time monitoring.

AutoRV automatically computes clinical indices with acceptable consistency versus manual measurements: the system achieved bias and standard deviation (std) of 1.09 ± 2.25 mm for free wall tricuspid annular plane systolic excursion (TAPSE_{fw}) and 0.19 ± 6.64

AutoRV represents the first end-to-end system capable of automated RV assessment in TEE. Its high inference speed and ability to derive standard clinical parameters demonstrate its potential as a tool for continuous cardiac monitoring in critical care environments.

Advisor:
Prof. Emiliano Votta
Prof. Gabriel Hanssen Kiss

Co-advisors:
Maïlys Hau
Jinyang Yu

Academic year:
2025-2026

Key-words: Right Ventricle, Transesophageal Echocardiography, TAPSE, RVFAC, RV Strain, Deep Learning, Automated Monitoring, Intensive Care, Ultrasounds.

1. Introduction

1.1. Right ventriular failure in ICUs

The right ventricle (RV) is one of the four chambers of the heart and plays a critical role in blood circulation. It receives deoxygenated blood from the right atrium (RA) and pumps it into the pulmonary artery, directing it toward the lungs for oxygenation. It has a crescent shape and, from a functional standpont, its wall has two sides: the interventricular septum, which separates the right and left ventricle (LV) and is shared between them, and the free wall (FW), which, under physiological conditions, moves like a bellows to induce blood ejection into the pulmonary circulation (fig. 1). The RV is characterized by low resistance and high compliance. Thus, as compared to the LV, it operates at lower pressures and is characterized by a thinner FW, which in turn reflects into limited tolerance to pressure overload. Since the advent of echocardiography, the lower pressure levels have led to considering RV pathologies less significant than LV ones. Indeed, the RV is often referred to as the "neglected" or "forgotten" chamber. Its retrosternal location, along with its complex structure and shape, has contributed to this perception [1].
Nonetheless, respiratory and circulatory failure account for a substantial portion of ICU admissions. Many of these patients require invasive mechanical ventilation, and a significant percentage, that varies from 20% to 50% depending on the underlying condition, develop RV dysfunction [2–4].
Mechanical ventilation functions by applying positive pressure at the airway level, at least during tidal ventilation and sometimes during expiration. This leads to pulmonary capillary compression and to increase in pulmonary vascular resistance and pleural pressure, raising RA pressure, reducing cardiac output, and increasing RV afterload [5–8].
Consequently, the RV tends to dilate, as this is its primary adaptive response due to high diastolic compliance. This dilation, known also as *acute cor pulmonale*, is dangerous as it leads to increased filling pressures and systemic congestion, and it can compress the LV, reducing the cardiac output and systemic arterial pressure, potentially causing circulatory shock [5, 7].
Usually, a first adaptation to this condition is the enlargement of the tricuspid valve (TV). This enlargement produces acute regurgitation, preventing further dilation at the cost of increased congestion and reduced forward flow. The reduced arterial pressure and contraction asynchrony further aggravate the condition by causing RV ischemia [7]. Acute RV failure arises from many causes, most commonly pulmonary hypertension (PH). It can develop after sudden increases in pulmonary pressure, and even with mild pressure elevation during conditions such as acute respiratory distress syndrome (ARDS), sepsis, or LV failure [5, 7, 9, 10].
RV failure can develop unpredictably, even in a previously healthy heart, or in a subtle manner that progressively worsens the patient's condition [11]. Indeed, the cooperation of all these contributing causes can lead to sudden death of the patient if left untreated [5–7, 9, 12, 13]. Thirty-day mortality after non-cardiac, non-obstetric surgery in patients with PH appears to range from 2% to 18% for elective procedures, rising to 15–50% in emergency operations [12]. Therefore, the early detection of RV disfunction and continuous monitoring in ICUs is of utmost importance, as prompt treatment, aimed at improving RV preload, contractility and afterload, is vital and precautions must be taken to prevent further injury [6, 12, 13].
Historically, assessment of RV function in intensive care has relied on the real-time and invasive measurement of hemodynamic data in the pulmonary artery by means of a pulmonary arterial catheter (PAC). However, this procedure is risky and has gradually been replaced by critical care echocardiography, which can yield more detailed information [6, 7, 10, 12, 13]. As a result, contemporary ICUs rarely employ PACs in standard practice [5, 6, 11].

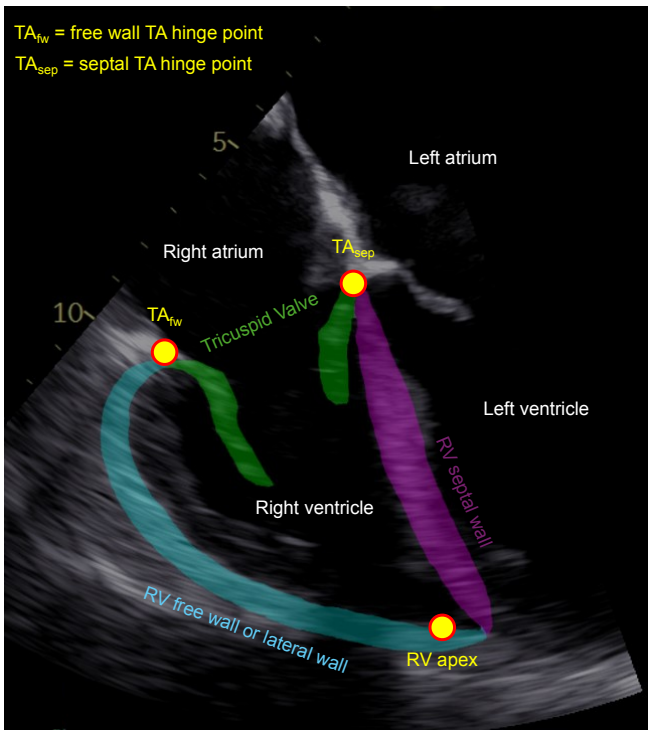


Figure 1: Main anatomical components in the right ventricle as visible in TEE imaging: free wall (light blue), septum (purple) and tricuspid valve (green). The two points on the tricuspid annulus (TA_{fw} and TA_{sep}) and the ventricular apex (RV apex) are highlighted in yellow.

1.2. Echocardiography

Echocardiography, an imaging technique developed in the 1960s, is the most commonly performed cardiac examination after electrocardiography and chest X-ray. Based on ultrasounds, echocardiography uses sound waves directed at the heart and reflected back to produce images. It includes various modalities, depending on how the ultrasound beam is steered and where the probe is positioned to acquire the signal.
The first distinction refers to different acquisition techniques, the main ones being M-mode, B-mode, Doppler, and three-dimensional echocardiography. M-mode (motion mode) echocardiography uses a single, focused beam to produce a one-dimensional “ice-pick” view of the heart. It continuosly records the movements of the cardiac structures along a single dimension. offering excellent temporal resolution [14]. Although limited in spatial resolution, it remains a great tool for quantitative measurements.
B-mode (brightness mode or 2D) echocardiography uses the same principle, and creates real-time, two-dimensional, cross-sectional images of the heart. They are obtained by sweeping the ultrasound beam through an arc, allowing for the creation of fan shaped images, composed by multiple "lines". obtained with the same principles as for the M-mode [14]. It is the routine modality for echocardiographic assessment of cardiac structures.
Doppler echocardiography measures the frequency shift (Doppler effect) of ultrasound waves caused by the motion of the cardiac structures, especially red blood cells within the circulation. Doppler echocardiography is widely used to evaluate valvular function, detect and quantify regurgitation or stenosis, and assess overall flow patterns within the heart and great vessels. It is an essential tool for studying both systolic and diastolic function and for estimating pulmonary artery pressures in clinical practice [14].
Three-dimensional echocardiography extends B-mode techniques to produce volumetric images. This is achieved with a matrix array probe that is composed of thousands of piezoelectric elements and can steer the scan line in lateral, elevation and azimuth planes, generating a detailed 3D image [15]. This is particularly useful to have better spatial visualization, to simplify communication between physicians, allowing a better orientation during surgeries. It can be used to measure volumes and blood flows, but at expense of temporal resolution, higher equipment expenses, and a steeper learning curve for image acquisition and interpretation [16–18]. For these reasons, 2D echocardiography is still the primary method thanks to its accessibility and speed.
The second distinction refers to different probe positions. In TTE, the probe is placed on the chest wall, and the ultrasound beam is transmitted through the thorax. It’s the most common echocardiographic technique, because it is non-invasive, and it is easier to use. However, image quality can be hindered by obesity or lung interference, and TEE is preferred in ICUs.
In TEE, the probe is mounted on the tip of an endoscope and inserted into the esophagus, allowing it to be

placed in close proximity to the heart. TEE is especially useful during surgery, as the probe can remain in place while monitoring cardiac performance throughout the procedure, guiding surgical decisions [14]. TEE can be both 2D and 3D. 2D TEE has a higher temporal resolution than 3D TEE but the former one is more intuitive, it avoids the foreshortening problem and the variability in tomographic planes, and facilitates the calculation of some useful indices based on volumes, like the right ventricular ejection fraction (RVEF) [16, 19].
In summary, echocardiography is non-invasive, efficient and accessible, and provides extensive information about RV morphology, dimension, contractility and hemodynamics [10, 12–14].

1.3. Echocardiographic indexes to evaluate the RV function

However, echocardiographic assessment and monitoring of RV function during management can be difficult. Usually, RV function is evaluated by qualitatively "eyeballing" it on 2D images [9]. Although this approach is fast, it remains qualitative, which could neglect quantitative changes that happen during management. It also lacks standardization, and evaluating function remains challenging due to the difficulty in accurately depicting the RV complex shape [20–23]. For this reason, many indexes have been introduced to easily and objectively measure the RV functionality.

1.3.1 Tricuspid annular plane systolic excursion

Tricuspid annular plane systolic excursion (TAPSE) is defined as the excursion of the lateral TA toward the RV apex from end-diastole to end-systole (fig. 2). It reflects the longitudinal shortening of the RV FW, therefore it serves as an indicator of the functional integrity of these structures [24]. It is a simple yet effective index, most commonly measured using TTE in M-mode from the apical window, and has been employed for decades to assess RV systolic function [1, 24]. Because it requires no geometric assumptions and correlates closely with RVEF, TAPSE is considered a reproducible, practical, and robust parameter of RV performance [24–28].
Particularly, TAPSE provides valuable prognostic information in patients with pulmonary arterial hypertension (PAH) and idiopathic PAH, and a low value is associated with RV dysfunction [19, 26, 27].
TAPSE can be assessed using either TTE or TEE; however, the clinical significance of TEE-derived TAPSE remains uncertain [29]. Furthermore, although TAPSE can be measured from multiple echocardiographic views and using different methods in both TTE and TEE, the resulting values are not necessarily interchangeable, as even minor variations in imaging planes can lead to statistically significant differences [30]. Consequently, no universally accepted cutoff value for abnormal RV function has been established, although several studies have reported threshold ranges between 15 and 18 mm [17, 20, 27, 31, 32]. Moreover, the underlying assumption that the displacement of a single segment reflects global mechanics is an oversimplification. Finally, the metric is angle dependent even when derived by the same view, and influenced by loading conditions [1, 33]. Despite these limitations, TAPSE remains a practical, reproducible, and clinically meaningful measure of RV systolic function, applicable in a wide variety of pathological contexts involving right heart dysfunction. In our study, some variations of the TAPSE index are measured in TEE acquisitions. These include TAPSE_{fw}, septal tricuspid annular plane systolic excursion (TAPSE_{sep}), and global tricuspid annular plane systolic excursion (TAPSE_{global}). Their definition is explained in section 2.2.

1.3.2 RV strain (RVS)

RV strain is a dimensionless metric that quantifies the percentage of systolic shortening of the right ventricular FW from base to apex [34], and it is measured as reported in fig. 2. While the standard RV FW strain uses only the three segments of the FW, a more complete evaluation of the RV can be obtained by a six-segment model, that averages strain from both the FW and the interventricular septum [35]. While these measurements have higher prognostic value over traditional indices, and are better predictors of RV dysfunction and failure [1, 19], they are less feasible, and rely heavily on the quality of the acquisitions [34, 36].

1.3.3 RV longitudinal shortening fraction

While TAPSE is easy to calculate, and provides strong prognostic value in assessing RV function, it has several notable limitations. Indeed, it relies on the assumption that the motion of a single segment reflects global mechanics, it is angle dependent, its normal range remains debated, it varies with loading conditions, and its interpretation depends on RV size [1, 33, 37].
Indexing TAPSE to RV lenght (RVL) enhance its predictive accuracy and alignment with RV FW strain and other RVL-independent indices, while having greater feasibility than RV FW strain (97.2% feasibility compared to 78,4%) [36]. Dividing TAPSE by the end-diastolic RV lenght (ED-RVL) improves its predictive performance

without adding much complexity to the calculation, as the cardiologist still measures TAPSE as usual and only needs to additionally measure the ED-RVL during acquisition [38].
Slightly different from this, RV longitudinal shortening fraction (RVLSF) measures the percentage change in the longitudinal dimension of the RV from diastole to systole (fig. 2). It is calculated as:

$$RVLSF = 100 \times \frac{DRVL_{mid} - SRVL_{mid}}{DRVL_{mid}}$$

where: $\mathrm{DRVL_{mid}}$ = midannular end-diastolic RV length, $\mathrm{SRVL_{mid}}$ = midannular end-systolic RV length,

RVLSF is not related to height, weight, and body surface area. People with high RVLSF ($\geq 16.5\%$) seem to have better survival than patients with low RVLSF ($\leq 14\%$). RVLSF has a predictive value for adverse outcomes comparable to that of RV FW strain (area under the curve (AUC) 0.82 vs. 0.84), despite the latter is more complex to calculate. This value is also much higher than that of TAPSE, which had an AUC of 0.55 [33, 37]. Given these findings, and because RVLSF correlates to RVS measured by speckle tracking echocardiography (STE) and to RVEF measured by cardiac magnetic resonance (CMR) [33, 37], RVLSF will be called RV linear strain (RVLS) from now on, and some variants will be proposed in this study, e.g., RV linear free wall strain (RVLS_{fw}), RV linear septal strain (RVLS_{sep}), RV linear midpoint strain (RVLS_{mid}), RV linear global strain (RVLS_{global}) (section 2.2). RVLS could be a simple but very effective way to assess RV function independently of physical characteristics of the patient.

1.3.4 RV fractional area change

Both TAPSE and RVLSF fail in evaluating RV contractility in the transversal direction. Although in genereal TAPSE strongly correlates with RVEF, suggesting the greater importance of the longitudinal shortening in evaluating RV function [28], in patients with increased RV afterload and RV dilation the transversal component increases in importance. This highlights the need for an index that more comprehensively reflects RVEF while remaining accessible and minimally invasive. Such an index consists in RV fractional area change (RVFAC). Based on the tracing of the RV endocardial surface at end diastole (ED) and end systole (ES) in a four-chamber view, the silhouetted areas RV end-diastolic area (RVEDA) and RV end-systolic area (RVESA) are obtained, and RVFAC is computed as:

$$100 \times \frac{(RVEDA - RVESA)}{RVEDA}.$$

As compared to TAPSE and RVLSF, RVFAC is more time-consuming to quantify and less reproducible [19, 26], but it is still relatively simple to calculate and shows a stronger correlation with cardiac magnetic resonance-derived ejection fraction [28]. A RVFAC $< 35\%$ indicates RV systolic dysfunction and has been associated with increased risk of death [19, 20]. Notably, RVFAC maintains its correlation with RVEF even in the later stages of disease, when TAPSE and RVLSF may lose predictive accuracy.

1.3.5 Tricuspid annular diameter

Although not directly related to RV contractility, TA diameter (TAD) is an indicator of TA dilation, and it is a predictor for postoperative RV dysfunction in patients undergoing tricuspid valve surgery [19]. Its measurement shows moderate agreement between TTE and TEE, which is potentially useful for mechanically ventilated patients in ICUs [29]. When leveraging on 2D TTE, TAD should be measured in the apical four-chamber view, which yields the highest feasibility and the best reproducibility among TEE views, reliably capturing the degree of enlargement albeit underestimating TAD as compared to 3D measures [39]. In synthesis, TAD emerges as a meaningful morphologic index that complements and integrates information obtained from TAPSE, RVS, and RVFAC.

1.3.6 Final remarks on monitoring by echocardiography

Taken together, these indices capture complementary aspects of RV performance and offer a more complete evaluation of RV function than any single parameter alone. Unfortunately, quantitative monitoring of the RV from TEE is time consuming and requires off-line analysis. Analyzing different indexes may require different imaging modalities, further complicating this approach [19, 40]. Furthermore, all these monitoring modalities require the presence of an expert clinician, diverting cardiologists and anesthesiologists from their main duties, which also leads to a higher cost [41, 42].
An echocardiographic modality developed to partially automate RV assessment is STE. This technique, introduced almost two decades ago, tracks the displacement of speckles, acoustic markers generated by the interaction between the ultrasound beam and myocardial tissue, throughout the cardiac cycle, using dedicated software [42, 43]. Initially applied to the LV, it has since been applied to other chambers, including the RV, where it

supports the detection of functional impairment through RVS [42, 44]. As a semi-automated method, it reduces inter- and intra-operator variability and shortens analysis time, and recent advances have extended it to 3D imaging, though with lower performance than 2D [42]. However, STE remains limited by the need for high-quality images and adequate frame rates, as well as by the requirement for trained operators and time-consuming offline analysis, which constrains its feasibility in routine practice [42].
These issues could potentially be solved by means of artificial intelligence (AI), particularly by training machine learning (ML) models, based for example on convolutional neural networks (CNNs) or vision transformers (ViTs), in order to predict the indexes and monitor RV function in real time and automatically.

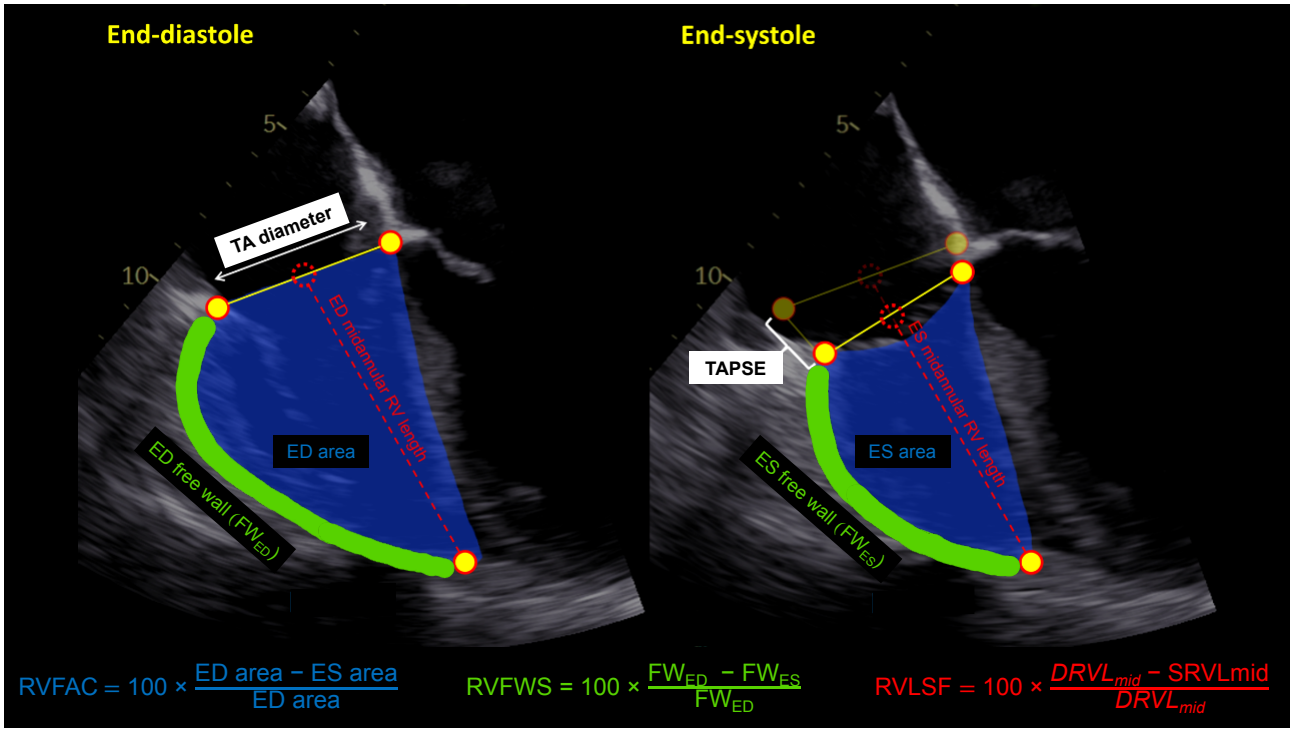


Figure 2: Illustration of how the principal indices for RV evaluation should be computed. The free wall length is shown in green, and the TA diameter in yellow, both at ED and ES. The RV areas at ED and ES are outlined in blue, while the corresponding midannular RVL is represented by a red dashed line. The measure of TAPSE for this example is also reported. The formulas of RVFAC, RVS and RVLSF are reported at the bottom of the figure, with colors corresponding to the measurements appearing in each formula.

1.4. Artificial intelligence to automate RV evaluation

ML is a subfield of AI that focuses on building models that learn directly from data, without being specifically programmed for a task. These models work by finding patterns in the data, and make predictions based on those patterns. Specifically, supervised learning is based on couples of inputs (X_n) and outputs (Y_n), and it has the objective of extracting a general function that predicts correct outputs for never seen inputs.
Traditional ML workflows rely on a human in the loop to extract meaningful features and often reduce data dimensionality using methods such as principal component analysis (PCA). This requires substantial domain expertise, as feature selection is typically based on intuitive, experience-driven judgments rather than explicitly defined data characteristics. For instance, distinguishing a dog from a cat may rely on abstract cues like ear shape or head conformation, yet translating such high-level concepts into pixel-level descriptors is extremely challenging, illustrating the inherent difficulty of manual feature engineering.
This led to the development of deep learning (DL). DL is a subfield of ML that uses more complex models, usually very deep and heavy neural networks, to process high dimensional, unstructured data like images, videos, audio or text. Since these models are so complex, they do not need manual feature extraction, as they inherently integrate this step inside their architecture. This means that they are free to select the best combination of features on their own. The costs for this improvement are computational complexity, training and inference time, and difficult interpretability. Indeed, since DL models are not based on a set of selected features, it is often very difficult to tell what features lead to the model decision.
CNNs are among the most widely used DL architectures for computer vision. Inspired by the organization of the

visual cortex, CNNs belong to the family of artificial neural networks (ANNs) but operate using local receptive fields. Their core component is the convolutional block (figs. 3 and 4), which applies learnable $n \times n \times \#channels$ filters across the spatial dimensions of the input to generate feature maps that capture relevant patterns. These blocks typically include nonlinear activation functions (e.g., rectified linear unit (ReLU)), normalization layers, and pooling operations to reduce spatial resolution. By stacking multiple blocks, CNN learn increasingly more abstract representations: shallow filters encode fine details, whereas deeper filters capture broader, more general structures. At the end of this dimensionality reduction, the feature maps are usually flattened and passed through a dense layer, that learns to perform tasks such as classification or regression.
CNNs are known since the 90's, but their popularity and development exploded in 2012, when AlexNet [45] paper was released. Since then, they have been the most popular, and for a long time the best performing algorithms in computer vision tasks, thanks also to their versatility across numerous different tasks. This popularity led to many studies being published, and many famous CNN architectures were developed in the following years.

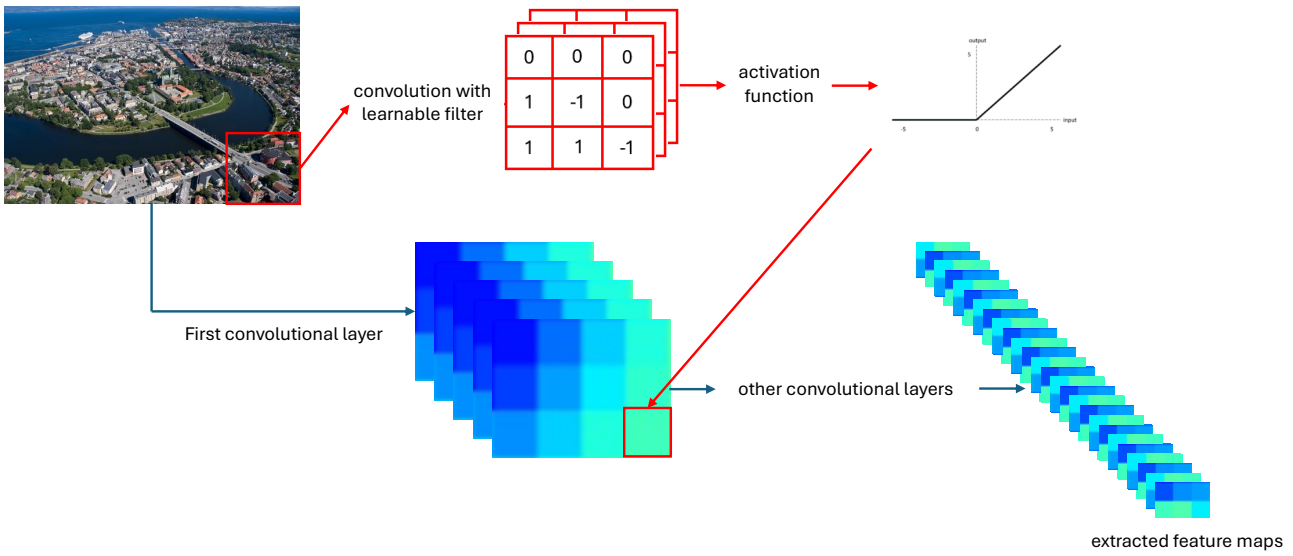


Figure 3: A schematic representation of a convolutional block. The feature maps from the previous layer, or the original image in the first layer (left), are convolved with learnable filters. The resulting feature maps are passed through an activation function, after which additional operations, such as normalization or pooling, may be applied. The output constitutes the feature maps of the subsequent layer. Convolutional blocks are typically stacked to form CNNs (fig. 4).

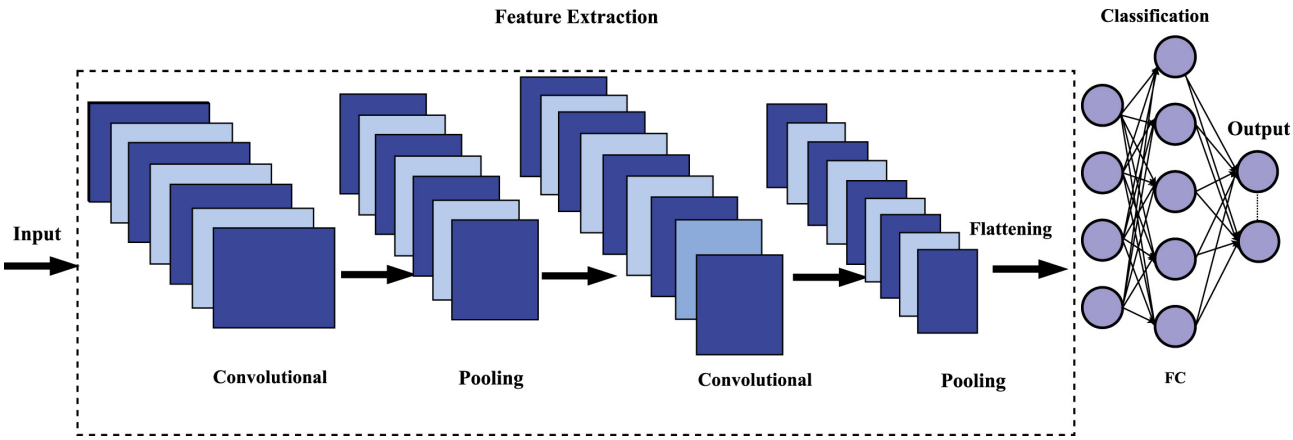


Figure 4: Representation of how convolutional blocks are stacked together to form the feature extractor of CNNs. The most important features are selected by the feature extractor, and are usually handed to a "head", that is specific to each task. Taken from [46]

1.4.1 Milestone CNN architectures

AlexNet (2012) [45] Built for ImageNet image classification competition, this CNN was composed of 5 convolutional layers (filter size 11x11, 5x5, 3x3, 3x3, 3x3) and 3 dense layers (2048, 2048, 1000). It was the first to use max-pooling instead of average pooling after the convolutions, and reached a performance that was the state of the art (SOTA) for the time.

ResNet (2015) [47] Several conceptual developments in CNNs led to the introduction of ResNet. The work in [48] showed that deeper networks with smaller filters are theoretically advantageous: stacking multiple 3×3 filters achieves the same receptive field as a single 7×7 filter, while using fewer parameters and introducing more nonlinearities.
However, simply increasing depth did not consistently improve performance, as deeper models often underperformed their shallower counterparts. This degradation was due to the difficulty of learning identity mappings, which in principle should allow a deeper model to reproduce a shallower one. The problem was worsened by vanishing gradients, where gradients diminish rapidly through many layers due to the multiplication of many near-zero terms in the backpropagation formula, preventing effective training of early layers.
To enable deeper architectures, He et al. [47] introduced residual connections (fig. 5), which bypass certain weight layers with identity mappings. These shortcuts make identity transformations easy to learn and provide more direct gradient flow, substantially mitigating vanishing gradients and enabling the successful training of very deep networks. ResNet architectures are commonly implemented at several depths, including 18, 34, 50, 101, and 152 layers

ResNeXt [49] Inspired by ResNet [47] and the inception module [50], in 2016 Xie et al. developed ResNeXt [49]. The basic architecture is the same as for ResNet, but the typical ResNet blocks are substituted with split-transform-merge blocks (fig. 5). The capacity of the model is increased by increasing the cardinality, described as the number of parallel transformations aggregated in each block. It was empirically proven that increasing cardinality leads to better accuracy than to increase width or depth. The output of the aggregated transformations is then summed together and the identity transformation is added as for the normal ResNet architecture.

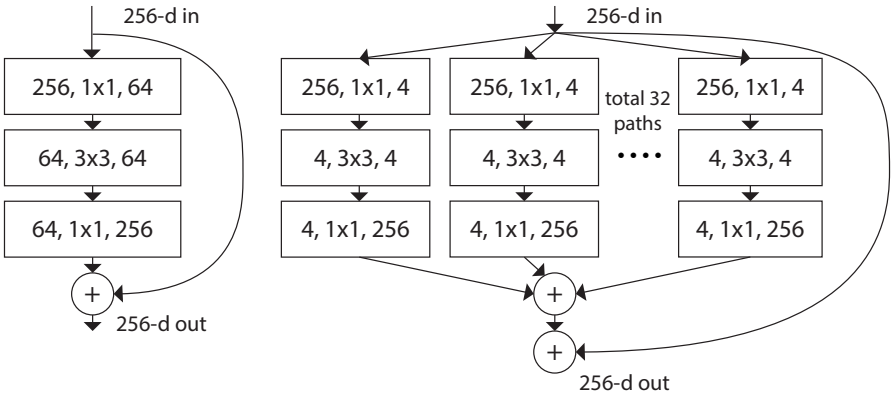


Figure 5: Comparison of two neural network building blocks. The left diagram shows a standard residual block [47], where features pass through a single sequence of convolutions and are added back to the input. The right diagram shows a ResNeXt-style block [49], where the input is split into 32 parallel paths that process features separately and are then combined, allowing the network to learn richer representations while keeping the same input and output size. Taken from [49].

U-Net [51] Each of the previously described networks contained one or more final dense layers. This architectural choice makes them useful for tasks like image classification or regression, but is unpractical when dealing with other tasks, for example with semantic segmentation, where an output class needs to be predicted for each pixel. U-Net is a fully convolutional model, designed to have a symmetric encoder-decoder architecture (fig. 6). The contracting path reduces the dimension of the input image by using 3x3 convolutions and ReLU, and 2x2 max pooling to extract important features. In contrast, the expanding path uses 2x2 up-convolutions, followed by 3x3 convolutions and ReLU to bring the feature map dimension back to the original image size. Skip connections are used between layers at the same level to take high resolution maps from earlier layers too and concatenate with contextual rich information from deeper layers. This helps especially because it gives back the border information to the latest layers. Thanks to all of this, the feature maps coming from the last

convolutional layer are rich of both contextual and border information, and 1x1 convolutions return n feature maps, where n is the number of classes to be predicted. The actual segmentation masks are usually obtained by taking the maximum value along the channels to the output masks, so that each pixel is assigned to the most activated class. The U-Net architecture, originally introduced in [51], has evolved over the years through the incorporation of padded convolutions, dropout, residual units, and other enhancements. Nevertheless, the core design remains unchanged, and U-Net continues to represent a SOTA solution for fast and effective semantic segmentation.

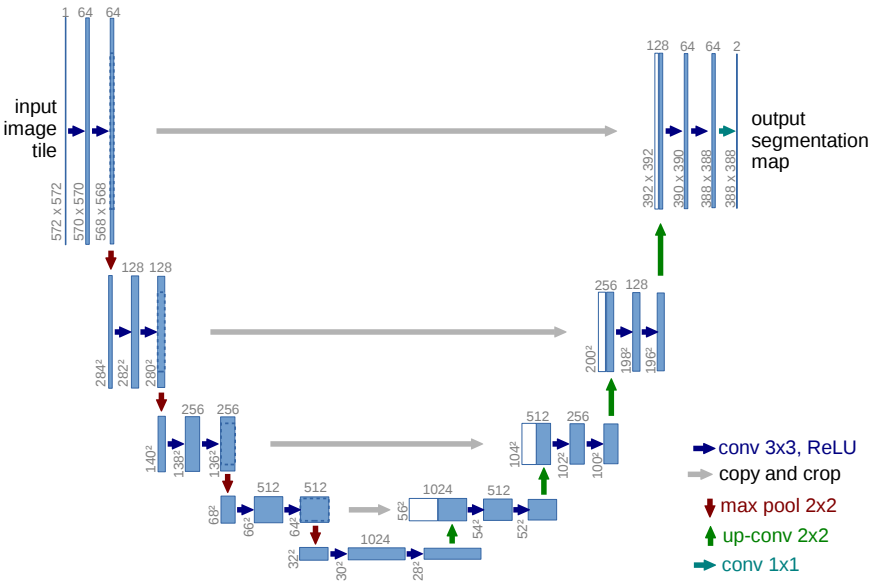


Figure 6: A representation of the U-Net architecture. The contracting path progressively reduces the spatial resolution of the input image while extracting high-level features, whereas the expanding path restores the spatial resolution to produce a segmentation label for each pixel. Taken from [51].

1.4.2 Transformer based networks

Natural language processing transformers Transformers are neural networks introduced for natural language processing (NLP) in 2017 [52]. Their key innovation is self-attention, a mechanism that computes context-aware representations of sequence elements. Each token is projected into queries, keys, and values, and attention weights are obtained through

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^\top}{\sqrt{d_k}}\right) V,$$

allowing the model to identify relationships within the sequence and refine token embeddings. The architecture consists of an encoder–decoder structure built from stacked self-attention and feedforward layers, with positional encodings added because attention lacks inherent order information. Transformers are highly parallelizable and achieved SOTA performance across NLP tasks.

Vision transformers Applying standard transformers directly to images is impractical because computing attention across all pixels scales quadratically with image size. Several works proposed approximations, such as restricting attention to local neighborhoods [53], using scalable global variants [54], operating on blocks [55], combining attention with CNNs [56], reducing image resolution [57], or dividing images into small patches [58]. A major breakthrough came with the ViT [59], which partitions the image into fixed 16×16 patches, flattens and linearly embeds them, and feeds them to a standard transformer encoder [52] as if they were tokens in a sequence. With large-scale pre-training, ViT achieved performance comparable to or exceeding CNNs in image classification.

Swin Transformer The Swin Transformer [60] was introduced to address the limitations of ViT [59] regarding fixed token scale and quadratic complexity. By employing variable patch sizes and local attention with shifted windows, the architecture achieves linear complexity relative to image size and enables cross-window connections. This hierarchical design facilitates dense tasks and integration with frameworks like FPN [61] or U-Net [51].

Swin UNETR Based on the same principles as the Swin transformer [60], in 2022 Hatamizadeh et al. [62] proposed a U-Net like network that incorporated a Swin-transformer as a feature extractor. The contracting part is composed by 4 stages, each one composed by 2 swin transformer blocks. Patch merging is applied to reduce the resolution by a factor of 2 at the end of each stage. The decoder is based instead on convolutions and on a classical U-Net design. Skip connections are applied as in the classical U-Net architecture between layers at the same level.

1.4.3 Current automatic pipelines for heart monitoring in echocardiography

As already stated, AI, and particularly DL, could be an effective way of monitoring the RV function in mechanically ventilated ICU patients. TEE is the preferred imaging modality in these cases, where the probe can be left in place and the RV can be accessed continuously. However, to the best of our knowledge, no solution currently exists for the automatic real-time monitoring of RV function based on TEE acquisitions. Indeed, most automatic pipelines focus on the LV, and the few ones that evaluate the RV do so by analyzing TTE acquisitions.

Left ventricular evaluation from TEE acquisitions Recent work has shown substantial progress in the automatic assessment of LV function from TEE images [41, 63–67]. Berg et al. [63] developed a CNN-based landmark detection method to estimate mitral annular plane systolic excursion (MAPSE) from 2D TEE, achieving high feasibility (>90%), fast processing (0.3 s per cycle), and low bias (0.5 mm with ECG, 0.2 mm without). In the same year, Tasken et al. [41] proposed a fully automated pipeline to analyze 3D TEE in real-time pipeline involving volume alignment, view detection, mitral annulus localization, and MAPSE estimation, reaching a mean difference and limits of agreement (LoA) of -0.16 ± 1.06 mm vs. manual measurement by expert clinicians. They later introduced Echocoder 2D+t [64], a 3D ResNet-34–based model predicting mitral annulus positions across 64 frames simultaneously (fig. 7), leveraging temporal information for improved accuracy. Finally, studies by Yu et al. (2024) [65, 67] confirmed the high feasibility and precision of automatic MAPSE estimation, with low bias and narrow limits of agreement compared to manual measurements.

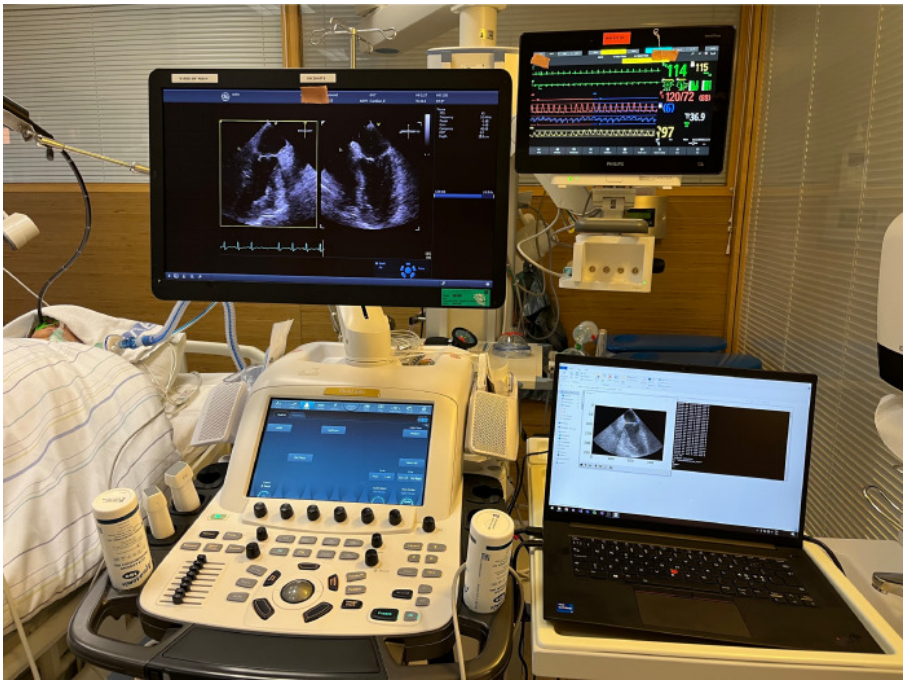


Figure 7: Application of Echocoder 2D+t in the ICU. A laptop provides the MAPSE measurement, together with visualizations, videos and plots. Taken from [64].

Right ventricular monitoring in TTE acquisitions Parallel to advancements in LV analysis, recent studies have demonstrated the efficacy of deep learning for RV assessment, although all of them use TTE. Beecy et al. [68] assessed RV function on 2D-TTE using a U-Net architecture with residual units [47, 51]. They segmented the TA in each frame and derived linear and circumferential TA displacement from the resulting masks. Their method achieved agreement with clinical measurements, with a reported bias and standard deviation of 0.15 ± 6.33 mm. Shad et al. [69] subsequently introduced a video-based AI system combining

spatial and optical flow streams from TTE into a 3D ResNet-152, predicting post-operative RV failure with an AUC of 0.729. Focusing on volumetric function, In 2023, Tokodi et al. [70] developed a fully automated pipeline to predict RVEF from 2D TTE. Their approach combined either a 2D ResNeXt [49] or ShuffleNet [71] backbone with a temporal convolutional layer and a regression head. Evaluated on an external dataset and compared with two expert operators, the system identified RV dysfunction with accuracy comparable to clinical readers (78.4% vs 77.0%), highlighting its diagnostic potential even with standard 2D imaging. Complementing this, Kampaktsis et al. [72] utilized a Transformer architecture on 2D TTE planimetry inputs, yielding high agreement with CMR reference standards. Most recently, Chernyshov et al. [17] introduced a real-time method for estimating TAPSE and RVFAC from 2D-TTE. Their approach relied on a light-weight U-Net–based pipeline [51] to automatically segment the RV wall, cavity, and RA. Contours were then extracted from the resulting masks, uniformly sampled with landmarks, and used to compute the clinical indices, achieving performance comparable to expert inter-observer variability (Bias and std of TAPSE = -0.4 mm ± 2.75 mm and of RVFAC = 0.8% ± 5.51%).

1.5. Objectives and approach

Based on all these considerations, this thesis project aimed to developing a fully automatic pipeline for RV monitoring from TEE images. The proposed approach is based on training DL models to detect clinically relevant landmarks in TEE midesophageal RV focused 2D acquisitions. The predictions are then fed to a simple algorithm that calculates clinically relevant indexes, such as TAPSE, RVS, and RVFAC, to quantify RV function. This approach was designed to obtain real-time inferences, i.e., to be suitable for bedside integration in ICUs, for aiding the clinicians in patient surveillance, and for signaling possible complications based on changes in the predicted parameters. The same motivation drove the choice of the different architectures evaluated in the study.

2. Materials and methods

2.1. Input and output

2.1.1 Clinical images

Mid-esophageal 2D TEE images focusing on the RV were obtained from mechanically ventilated patients under general anesthesia. An E95 scanner and a 6VT-D probe (GE Healthcare, Horten, Norway) were used in all patients. Scanner settings were adjusted to obtain a temporal resolution of at least 40 frames and, on average, each recording contained 2.94 heart cycles. As a result, the whole dataset consisted in 8424 frames.

2.1.2 Study population

This study included 52 patients scheduled for on-pump cardiac surgery. Their characteristics have been detailed in previous publications [64–66, 73]; however, the present analysis includes two additional patients. Exclusion criteria were age under 18 years, refusal to participate, or contraindications to TEE. The study was conducted according to the Declaration of Helsinki and approved by the Regional Committee for Ethics in Medicine (REK number 2022/345556). Written informed consent was obtained from all participants prior to inclusion. Table 1 summarizes the characteristics of the study cohort. Demographic data were unavailable for two of the original 52 patients, resulting in 96% completeness. All intraoperative RV measurements were obtained with a 3D TEE probe and computed over a single heartbeat.

2.1.3 Anatomical landmarks of interest

The quantification of RV geometry and function in this study relies on three landmarks identified at two key moments of the cardiac cycle, end diastole (ED) and end systole (ES) (fig. 8). The first two landmarks are the TA hinge points, located at the insertions of the TV leaflets on the RV free wall (TA_{fw}) and on the septal wall (TA_{sep}). Their positions at ED are denoted TA_{fw}^{ED} and TA_{sep}^{ED}, and at ES as TA_{fw}^{ES} and TA_{sep}^{ES}. The third landmark is the RV apex, defined at the junction between the relatively straight septal wall and the curved free wall, indicated as A^{ED} and A^{ES} at ED and ES, respectively.
Regardless of whether they are placed manually by a clinician or detected automatically with AutoRV, the ED and ES positions of these three landmarks form the basis for computing several clinical indices (see section 2.2).

Patient characteristics (n = 50)

Variable	Number of patients (percentage)
Male	38 (76%)
Preoperative SPAP 30 to 55 mmHg	9 (18%)
Preoperative SPAP $>$ 55 mmHg	4 (8%)
Preoperative LVEF $<$ 30%	4 (8%)
NYHA class 1	22 (44%)
NYHA class 2	14 (28%)
NYHA class 3	14 (28%)
NYHA class 4	0 (0%)
Diabetes mellitus on insulin	1 (2%)
Creatinine clearance $<$ 85 mL/min	20 (40%)
Variable	**Mean value ± std**
Age (years)	67 ± 8
Body surface area (m^2)	1.98 ± 0.2
Preoperative LVEF (%)	47 ± 12
Variable	**Median Value and interquartile range**
EuroScore II	1.8% [1.0% - 2.7%]
Intraoperative RVEF	43% [37% - 47%]
Intraoperative RVEDV	156 mL [129 mL - 182 mL]
Intraoperative RVESV	88 mL [74 mL - 108 mL]

Table 1: Summary table of the patients' demographic and characteristics. Intraoperative right ventricular measurements are made using single-beat three-dimensional transesophageal echocardiography. Data was missing in two (4%) of the patients. Abbreviations: LVEF, left ventricular ejection fraction, NYHA, New York Heart Association; RVEDV, right ventricular end-diastolic volume; RVEF, right ventricular ejection fraction; RVESV, right ventricular end-systolic volume; SPAP, systolic pulmonary arterial pressure.

2.2. Manual Quantification of Clinical Indexes

The RV morphological and functional indices listed in table 2 were manually quantified by expert clinicians using EchoPAC (GE Vingmed Ultrasound, Horten, Norway). For each distance measurement, the three landmarks of fig. 8 were re-traced from scratch at ED and ES, which were selected based on the expansion of the RV chamber. TAPSE was measured at the free wall (TAPSE$_{fw}$), at the septum (TAPSE$_{sep}$) and at TAPSE$_{global}$. TAPSE$_{fw}$ is computed as the distance between TA_{fw}^{ED} and TA_{fw}^{ES}, while TAPSE$_{sep}$ is the distance between TA_{sep}^{ED} and TA_{sep}^{ES} (fig. 8). TAPSE$_{global}$ is then calculated as the average of these two measures.
As shown in table 2, TA end-diastolic diameter (TADD) and TA end-systolic diameter (TASD), as well as RV septal and free wall lengths (septal ED-RVL (ED-RVL$_{sep}$), free wall ED-RVL (ED-RVL$_{fw}$), septal ES-RVL (ES-RVL$_{sep}$), and free wall ES-RVL (ES-RVL$_{fw}$)) lengths are measured for both diastole and systole using the Caliper tool in EchoPAC. Measures were obtained by placing a marker on the TA hinge points, and on the RV apex and measuring the distances (table 2 and fig. 8). Each measurement is done over three to five cardiac cycles before being averaged into one value. The RVL from the midpoint of the TA to the RV apex is then calculated for ED and ES (midpoint ED-RVL (ED-RVL$_{mid}$) and midpoint ES-RVL (ES-RVL$_{mid}$) respectively) with the general formula:

$$RVL_{mid} = \sqrt{\frac{1}{2} \times (RVL_{fw}^2 + RVL_{sep}^2) - \frac{1}{2} \times TA_{diameter}^2} \tag{1}$$

Furthermore, using the same three points, RV linear strain (RVLS) is calculated. This measurement was computed for the septal, free wall and RV midpoint lengths (see table 2) using the following generic formula:

$$RVS = 100 \times \frac{Diastolic\ \ Length\ -\ Systolic\ \ Length}{Diastolic\ \ Length} \tag{2}$$

Additionally, RV linear global strain is also computed using:

$$RVS_{glob} = 100 \times \frac{(EDRVL_{fw} + EDRVL_{sep}) - (ESRVL_{fw} + ESRVL_{sep})}{EDRVL_{fw} + EDRVL_{sep}} \tag{3}$$

Finally, RVFAC is calculated using RVEDA and RVESA, each measured by tracing the RV endocardium at ED and ES.
This yields several important parameters for the structure and function of the RV (see table 2 for an in-depth description of each acronym).

2.3. AutoRV

Our proposed AutoRV is an automated tool that performs two tasks: landmark detection to identify the TA hinge points and the RV apex (fig. 9), by means of a CNN, and computation of the morphological and functional indexes from identified landmarks. The underlying methods are described in the following subsections.

2.3.1 Manual Creation of Ground Truth

As mentioned in section 2.1.1, the dataset to train, validate, and evaluate the CNNs consisted of a total of 8424 frames, of which 7799 were manually annotated by a trained operator under the guidance of a clinician, who provided one annotated frame per recording and subsequently reviewed the annotations. Three landmarks were identified and traced: the two TA hinge points and the RV apex (fig. 9).
Upon visual inspection of the full cardiac cycle, ED and ES were identified as the frames with maximal and minimal base-to-apex RV dimension by the annotator. Two ranges spanning over at least seven frames, centered around ED and ES, respectively, were considered; within each range, every frame was manually annotated. In the remaining phases of the cardiac cycle, one frame out of two was annotated, and the position of the three landmarks in the remaining frames was obtained by linear interpolation of their positions at the preceding and subsequent frames. To take advantage of the temporal information while annotating, the operator focused on placing the same landmark on every considered frame, before restarting the video and identifying the next landmark.
At every relevant frame, TA hinge points were identified by observing the motion of the leaflets and by selecting the pivots of their movements at their insertion on the RV wall. When leaflets were poorly visible, the TA hinge points were identified based on their annotations at the preceding or subsequent frames, with the help of the local motion of the RV walls. Frames not containing all three landmarks were discarded; as a result, 7799 frames out of the initial 8424 were kept for the subsequent steps of the work.

Clinical indices used to evaluate our performances

Notation	Clinical relevance	Definition
Tricuspid annular plane systolic excursion		
TAPSE_{fw}	Established parameter of the RV function	$d(TA_{fw}^{ED}, TA_{fw}^{ES})$ see fig. 8
TAPSE_{sep}	Related to LV function since it is equivalent to septal MAPSE.	$d(TA_{sep}^{ED}, TA_{sep}^{ES})$ see fig. 8
TAPSE_{global}		$\frac{TAPSE_{fw}+TAPSE_{sep}}{2}$
Tricuspid annular diameter		
TADD	Reflect RV dimensions. Integrating a parameter of ED can be helpful because ED dimensions are proportional to preload.	$d(TA_{fw}^{ED}, TA_{sep}^{ED})$ see fig. 8
TASD		$d(TA_{fw}^{ES}, TA_{sep}^{ES})$ see fig. 8
RV areas		
RVEDA	Reflect RV dimensions. Likely more robust than TA diameters because they include more information. Necessary to calculate RVFAC.	Area delimited by the TADD and RV walls at ED.
RVESA		Area delimited by the TASD and RV walls at ES.
RVFAC	Established parameter of the RV function. Superior to TAPSE because it also reflects radial function	$100 \times \frac{RVDA-RVSA}{RVDA}$
End-diastolic RV lengths		
ED-RVL_{fw}	Necessary to calculate RVS.	$d(TA_{fw}^{ED}, A^{ED})$ see fig. 8
ED-RVL_{sep}		$d(TA_{sep}^{ED}, A^{ED})$ see fig. 8
ED-RVL_{mid}		see eq. (1)
End-systolic RV lengths		
ES-RVL_{fw}	Necessary to calculate RVS.	$d(TA_{fw}^{ES}, A^{ES})$ see fig. 8
ES-RVL_{sep}		$d(TA_{sep}^{ES}, A^{ES})$ see fig. 8
ES-RVL_{mid}		see eq. (1)
Linear RVS		
RVLS_{fw}	Analogous to RV free wall strain. More commonly reported than global or septal metrics of the RV.	see eq. (2)
RVLS_{sep}	Related to LV function. Necessary to calculate RVLS_{global}.	see eq. (2)
RVLS_{mid}	Highly correlated to strain by speckle tracking and RVEF by CMR [33, 37].	see eq. (2)
RVLS_{global}		see eq. (3)

Table 2: Description of a selection of clinical indices. $d(\cdot)$ denotes Euclidean distance between two points.

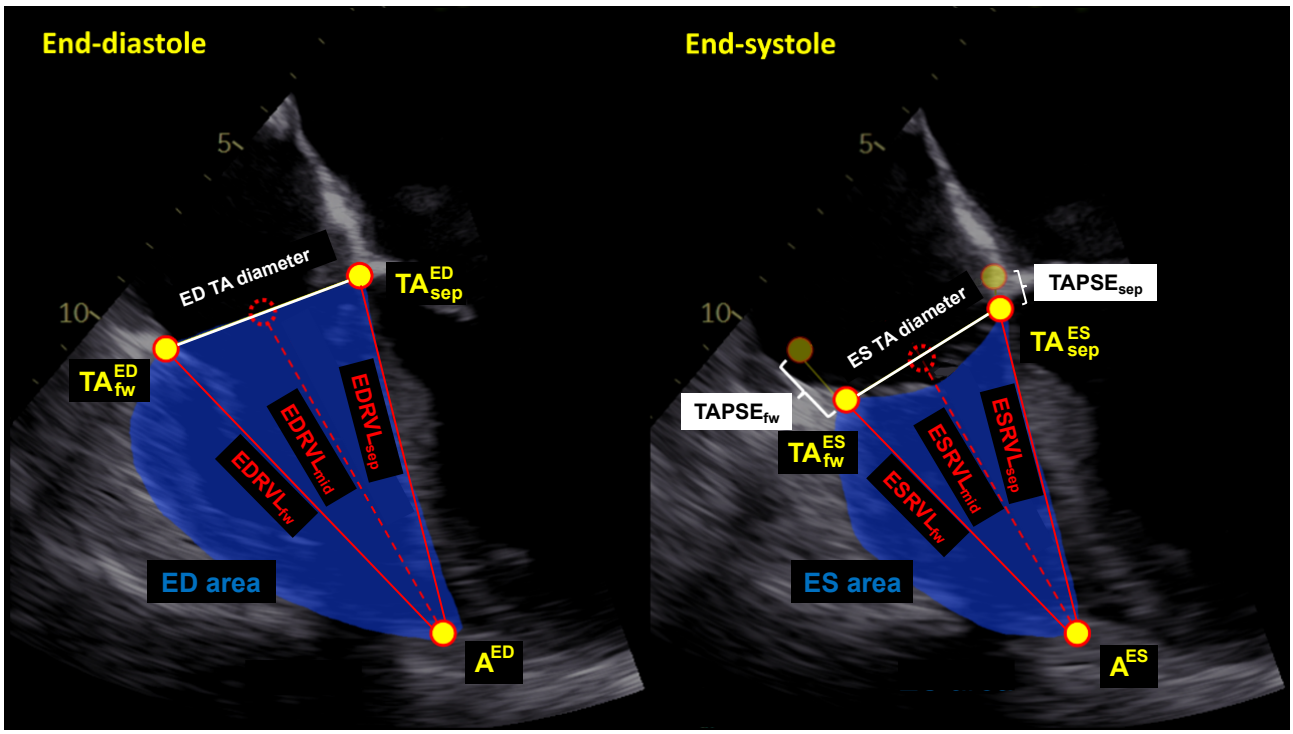


Figure 8: Main measures and landmarks considered in this study. The free-wall TA hinge point (TA^{ED}_{fw} and TA^{ES}_{fw}), the septal TA hinge point (TA^{ED}_{sep} and TA^{ES}_{sep}), and the RV apex (A^{ED} and A^{ES}) are colored in yellow. The distances between the free wall and septal TA landmarks at ED and ES are TAPSE_{fw} and TAPSE_{sep} respectively. The RV lengths of the free-wall ($EDRVL_{fw}$ and $EDRVL_{fw}$), of the septum ($EDRVL_{sep}$ and $EDRVL_{sep}$) and of the TA midpoint ($EDRVL_{mid}$ and $EDRVL_{mid}$) are represented in red at ED and ES. These were used to calculate RV linear strain. The ED and ES tricuspid annular (TA) diameters are traced in white. Finally, the ED and ES area are traced in blue, and are used to calculate RVFAC.

The dataset was randomly split patient-wise into training, validation, and testing sets, respectively 70%, 10%, and 20% (table 3). The acquisitions in the test set were selected among those with no discarded frames, as this was necessary to calculate the indices as expected in the real case scenario.

Detailed description of the dataset

	Recordings	Frames	Heart cycles				
			1	**2**	**3**	**4**	**5**
Train	36	5413	2	4	28	1	1
Validation	5	751		1	3		1
Test	10	1635		1	8		1
Total	**51**	**7799**	**2**	**6**	**39**	**1**	**3**

Table 3: Detailed description of our dataset across training, validation and test splits. When automatically detecting ED and ES, the method differs from that used by clinicians. Therefore, the number of detected heart cycles varies between the manual and automatic methods.

2.3.2 Automated Landmark Detection

We trained and tested six different CNN architectures: U-Net [51], three variations of ResNet [47], ResNeXt50 [49], and SwinUNETR [62]. ResNet, in all of its variants, and ResNeXt50 were trained to detect landmarks directly, in accordance with their original implementation. U-Net and SwinUNETR are, however, originally

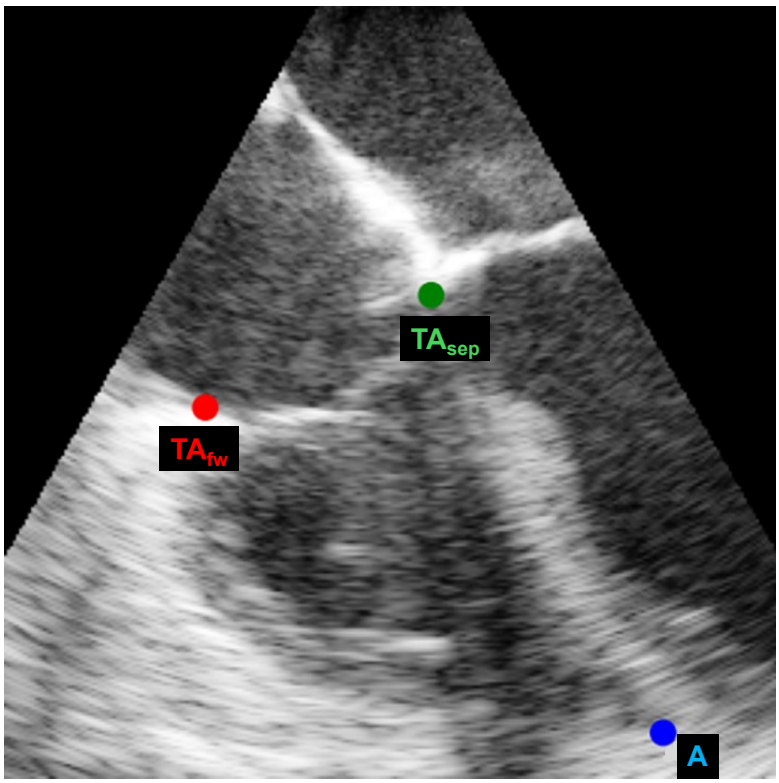


Figure 9: Example of ground truth annotation landmarks: in red and green, the TA hinge points (resp. free wall and septal), and in blue, the RV apex. Landmarks are detected for every frame, not just ED and ES.

implemented to segment images and not regress numbers; hence, we trained these to produce a heatmap for every landmark, which encoded at every pixel the image the probability for the landmarks to be located therein. The coordinates of each of the three points were then obtained by computing the barycenter of the top activated pixels above a threshold, with the threshold optimized during training as described in table 4. Before being fed to the CNNs, images were cropped and/or padded to 256×256 pixels, and a fixed lookup table was applied to enhance the input's contrast. To compensate for the relatively low numerosity of the dataset, data augmentation was applied through a combination of i) random Gaussian noise added with a σ of 0.04, ii) contrast and brightness deformation with a γ between 0.8 and 1.2, and iii) random rotations, with rotation angles falling in the [-15°, +15°] range, and random crops, with a minimum crop size of 230×230 pixels, and performed on a random selection of 60% of the inputs.
Training, validation, and testing were conducted using PyTorch 2.1 and MONAI [74] on an NVidia GeForce GTX 1080 graphics card with 8 GB of VRAM. For every architecture, the best set of hyperparameters was identified through preliminary experiments devoted to exploring the hyperparameter space. Specifically, the batch size, initial learning rate, number of input channels in U-Net, feature size of the SwinUNETR model, depth, dropout, number of residual units in each convolutional layer for the U-Net, and confidence thresholds described previously were explored and summarized in table 4.

Hyperparameters used during training

	U-Net	ResNet18	ResNet34	ResNet50	ResNeXt50	SwinUNETR
Batch size	128	16	16	16	0	16
Initial LR	[1e-4;1e-2], 5e-3	[1e-4;1e-2], 1e-3	[1e-4;1e-2], 5e-3	[1e-4;1e-2], 1e-3	1e-4	1e-4
# Channels	[4;32], 16	N/A	N/A	N/A	N/A	N/A
Feature size	N/A	N/A	N/A	N/A	N/A	12
Depth	6 down, 6 up	18	34	50	50	5 down, 5 up
Dropout	[0;0.25], 0	0	0	0	0	0
# Res. units	[0;2], 0	N/A	N/A	N/A	N/A	N/A
Threshold	[40%;5%], 12.5%	N/A	N/A	N/A	N/A	[30%;10%], 20%

Table 4: Hyperparameters used during training. When applicable, the tested intervals are reported in brackets, followed by the selected value.

As visible in table 4, the experiments were especially focused on the U-Net, since it produced better results from the beginning and had a faster inference speed, which made it preferable to architectures such as SwinUNETR. The learning rate was divided by a factor of 3 if the validation loss stopped decreasing for 3 epochs. The training was stopped if the validation loss stopped decreasing for more than 10 consecutive epochs. The model with the best validation loss was saved and used for further evaluation in section 3. The euclidean distance between predicted and target landmarks was used as a training loss. The performing model had a batch size of 128 and an initial learning rate of 0.005.

2.3.3 Automatic Calculation of Clinical Indexes

Based on the frame-by-frame predicted position of the landmarks (fig. 9), the clinical indices were automatically computed. To this aim, four approaches were implemented and tested:

1. computation based on the raw predicted coordinates of the three landmarks;
2. computation upon filtering the raw predicted coordinates of the three landmarks through a Kalman filter. The Kalman filter was initialized with a constant-velocity motion model, zero initial velocity, a process noise standard deviation of 5, and measurement noise standard deviation of 0.1. At each time step, the filter performed a prediction followed by an update using the observed coordinates, yielding the estimated filtered position.
3. computation upon filtering the raw predicted coordinates through moving linear interpolation: for a given landmark with predicted position $\mathbf{x}(t)$, and given three consecutive frames t_{i-1}, t_i, and t_{i+1}, if $||\mathbf{x}_t - \mathbf{x}_{t-1}|| > 2mm$ and $||\mathbf{x}_t - \mathbf{x}_{t+1}|| > 2mm$ then $\mathbf{x}(t_i)$ is set to the average of $\mathbf{x}_{t-1}$ and $\mathbf{x}_{t+1}$.
4. computation upon filtering the raw predicted coordinates with the moving linear interpolation first, and following with the Kalman filter.

The ED frame of each heartbeat was automatically identified based on the ECG signal recorded during the exam: R-peaks were detected using the Pan–Tompkins [75] algorithm, and ED frames were selected as the frames that are the closest in time to each R-peak. A heart cycle was then defined as the interval between two ED frames. Specifically, the ES frame was defined as the frame characterized by the maximal distance from $TA_{fw}(t)$ to the position of TA_{fw} at ED (TA_{fw}^{ED}). The TA_{fw} coordinates were specifically filtered for this procedure by applying the cascade of linear interpolation and Kalman filtering described above, in order to limit obvious errors as much as possible. TA_{fw} was considered instead of TA_{sep} because it was the most consistently predicted landmark when testing the model.

AutoRV computes all clinical indices from these two ED and ES frames, using the corresponding predicted landmark coordinates and applying whichever filtering strategy (none, either filter, or both) is most suitable for each specific index. The coordinates of the midpoint of the TA were calculated by averaging the coordinates of TA_{fw} and TA_{sep}. The predicted clinical indices are then calculated with the same formulas reported in table 2, with the exception of the RV areas.

Indeed, RV area at a given frame was calculated using a method designed to approximate manual endocardial tracings: TA_{fw}, A, and TA_{sep} were interpolated by a cubic spline, which was sampled at N equally spaced points; RV area was then computed as the area of the polygon defined by the N points. Based on a preliminary sensitivity analysis, N was set to 6 and 5 when computing RVEDA and RVESA, respectively, as these values provided the best agreement with manual measurements.

An unconventional strategy was adopted to reduce discrepancies in RVFAC vs. manual quantifications. These discrepancies were typically caused by the overestimation of RVESA, whereas RVEDA was correctly quanti-

fied. To compensate for the overestimation of RVESA, when computing RVFAC N was increased to 11 to compute RVEDA: this choice led to consistently overestimating RVEDA to an extent that compensated for overestimations of RVESA when computing RVFAC.
After computing each index per heartbeat across an acquisition, the final prediction was derived as the mean, median, maximum, or minimum value across all heartbeats. Finally, TAPSE_{global} was defined as the average of TAPSE_{fw} and TAPSE_{sep}, each filtered with its best-performing strategy and reduced using the method that individually yielded the best results. A detailed explanation of the filtering technique and reduction method across heartbeats is reported in table 5.

Clinical indices processing strategy

Group	Index family	Index	Filter	Reduction
TAPSE	TA motion	TAPSE_{fw}	Kalman	median
		TAPSE_{sep}	Both	maximum
		TAPSE_{global}	Strategy in section 2.3.3	
Diameters	TA distance	TADD	Kalman	median
		TASD	None	minimum
Areas	RV area	RVEDA	Both	maximum
		RVESA	Kalman	minimum
		RVFAC	Strategy in section 2.3.3	
DRVL	RV length	ED-RVL_{fw}	None	mean
		ED-RVL_{sep}	None	maximum
		ED-RVL_{mid}	None	median
SRVL	RV shortening	ES-RVL_{fw}	Kalman	minimum
		ES-RVL_{sep}	Both	minimum
		ES-RVL_{mid}	Both	minimum
RVS	Linear strain	RVLS_{fw}	Kalman	maximum
		RVLS_{sep}	Kalman	maximum
		RVLS_{mid}	Kalman	maximum
		RVLS_{global}	Kalman	maximum

Table 5: Filtering and reduction strategies used to compute each clinical index. "Filter" indicates the filtering strategy applied to the raw predictions, and “Reduction” indicates how the index was derived from multi-heartbeat measurements.

3. Results

3.1. Landmark detection

As mentioned in section 2.3.2, experiments were conducted using a U-Net and a SwinUNETR, that output three heatmaps, or various variations of ResNet. U-Net was the best performing architecture for validation loss by a significant margin, with an average distance error between predicted landmarks and ground truth of 4.42 mm (table 6). For this reason it was selected as the best performing model, and further analyzed on the test set.
The pipeline results transfer well from validation to testing set: the average error is similar, with a value ± std of 4.16±3.36 mm (table 7). However, this average value resulted from mismatches that were rather different from landmark to landmark (fig. 10 and table 7): 2.93 ± 1.80 and 3.52 ± 3.38 mm for TA_{fw} and TA_{sep}, and 6.04 ± 3.72 mm for A. The detection of A was hence much less reliable than the detection of the TA hinge points. By analyzing the predicted and ground truth positions for a set of randomly selected images, it was

Validation performance

Network	Distance (mm)
U-Net	**4.42**
ResNet18	6.01
ResNet34	6.92
ResNet50	6.33
ResNeXt50	6.35
SwinUNETR	5.13

Table 6: Best average landmark distance error on the validation set for the tested models. Best result shown in bold.

possible to observe that the predicted position of A could vary from points on the endocardial surface to points several millimeters inside the heart wall (fig. 11).

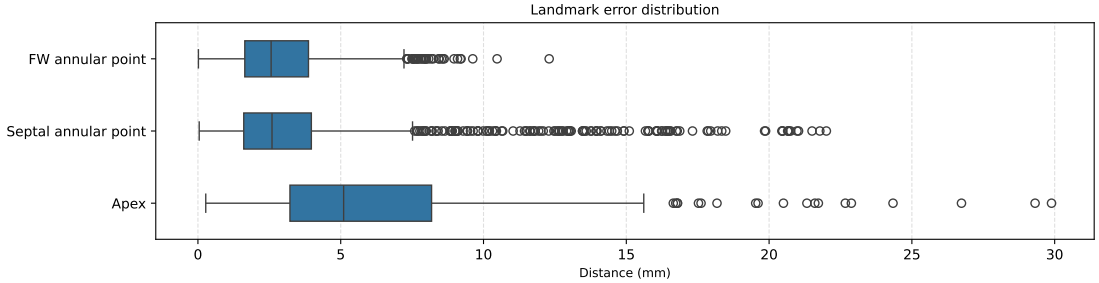


Figure 10: Box plot of distance errors between ground truth and predicted landmarks over the test set. Blue boxes indicate the interquartile range of the error distribution, with the central black line representing the median. Whiskers extend to the most extreme values within ($1.5 \times$ IQR), and points beyond this range are shown as outliers (empty circles).

Test set landmark errors

Free wall	Septal	Apex	Global
2.93 ± 1.80 mm	3.52 ± 3.38 mm	6.04 ± 3.72 mm	4.16 ± 3.36 mm

Table 7: Distance errors $\pm$ std between ground truth and predicted landmarks over the test set.

Moreover, the performance of the pipeline varied also from patient to patient, as some of those images were more challenging to analyze (fig. 12).
The analyzed U-Net model was also one of the fastest models, with an inference rate of 248 fps; well above the frequency of TEE acquisitions (table 8). This evidence, togheter with the errors obtained in the validation and test sets, which were almost always acceptable, indicated the U-Net model, among the considered solutions, as ideal for monitoring use in ICUs.

3.2. Clinical Indexes

Mismatches in the values of the clinical indices between AutoRV predictions and ground truth are summarized in table 9 and graphically represented by Bland–Altman plots (figs. 13 to 17). In general, good agreement was observed: for most patients, the differences between manual and automatically computed measurements were centered close to zero, with at most a single patient lying outside the LoA for each index.
Mismatches in TAPSE_{fw} (1.09 ± 2.25 mm) compare favorably with those obtained using the semi-automated

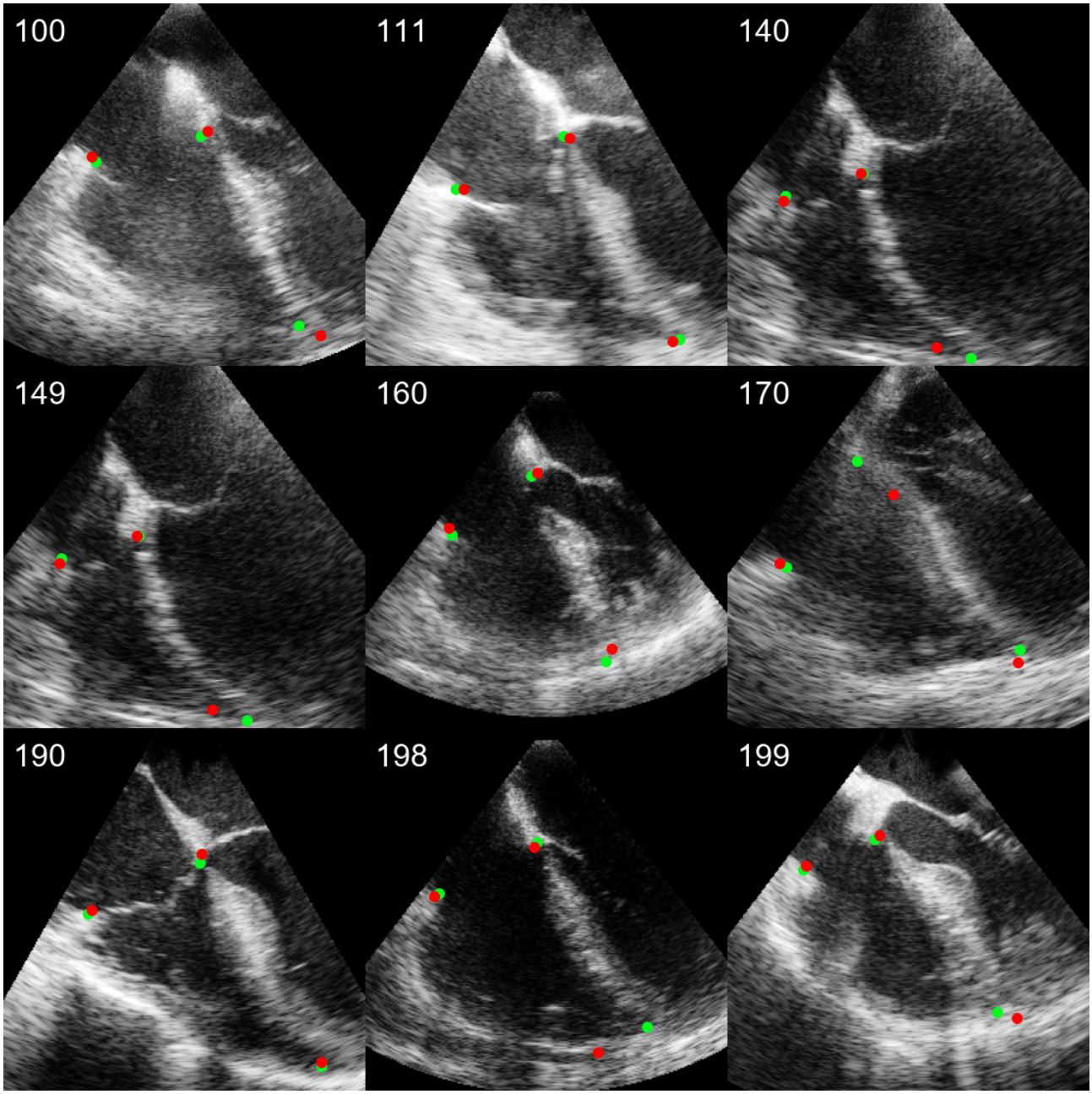


Figure 11: Random selection of example results for landmark detection by AutoRV in the test set. Green = ground truth; red = prediction. Images are identified by their patent ID.

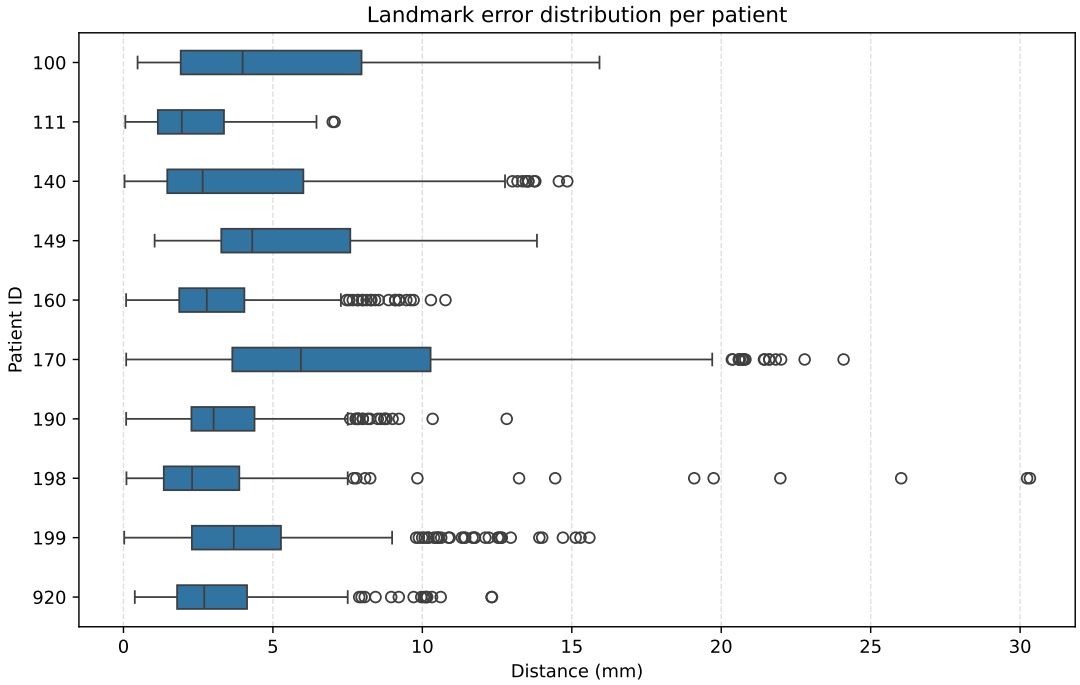


Figure 12: Box plot of average error per patient between ground truth and predicted landmarks over the test set. Blue boxes indicate the interquartile range of the error distribution, with the central black line representing the median. Whiskers extend to the most extreme values within (1.5 × IQR), and points beyond this range are shown as outliers (empty circles).

pipeline proposed by [17], which reported an error of -0.40 ± 2.75 mm. While our approach shows a slightly higher bias, it achieves narrower LoA, despite being fully automated and not requiring manual selection of ES and ED frames prior to index computation. Errors in TAPSE$_{fw}$ remain above the inter-observer variability reported in the literature [76, 77]. In particular, both the bias and std exceed the inter-observer variability of -0.04 ± 0.54 mm and the intra-observer variability of -0.52 ± 0.69 mm reported for lateral TA motion in the mid esophageal 4 chambers view (ME4C) view [76]. Furthermore, the normalized LoA of 47.29% is higher than the 9.7% reported for inter-observer variability by Skinner et al. [77].

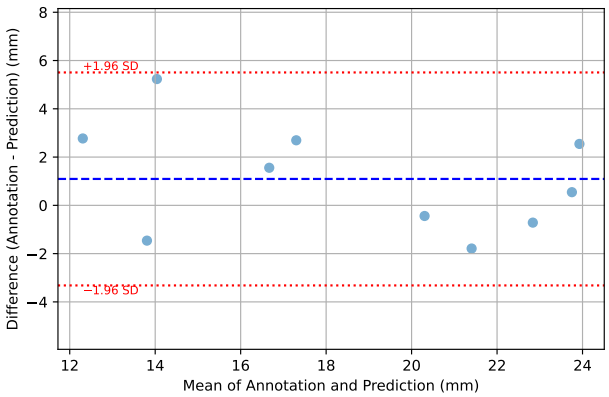
Indices incorporating the septal annular point, such as TAPSE$_{sep}$, TAPSE$_{global}$, TADs, and RVLs, tend to be less stable than those based solely on TA_{fw}. The performance of TAPSE$_{sep}$ is markedly inferior to that of the other TAPSE variants, with a bias and std of 0.93 ± 3.87 mm. Although the bias is lower, the substantially higher std leads to wider LoA. Given the smaller average value of TAPSE$_{sep}$, these absolute errors correspond to larger relative discrepancies and frequent inconsistencies across cases. In contrast, TAPSE$_{global}$ yields the best results among the computed TAPSE variants, with a bias and std of 1.01 ± 2.03 mm, remaining comparable to TAPSE$_{fw}$ in terms of bias and std.

The TAD is predicted with relatively high consistency. Both TADD (0.16 ± 3.67 mm) and TASD (-1.63 ± 2.64 mm) exhibit low bias and std, with TASD being slightly overestimated but consistent across measurements.

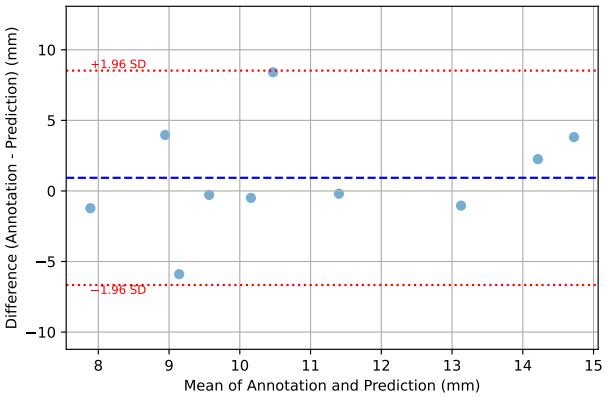
When considering summary statistics (table 9), RV areas appear well predicted by the proposed pipeline. Specifically, RVEDA showed a bias and std of -0.29 ± 3.73 cm^2, while RVESA was mildly overestimated, with 2.40 ± 2.74 cm^2.

Inspection of the Bland–Altman plots (fig. 15) and representative examples (fig. 11) revealed that patient 190 presents a particularly thin RV morphology, associated with large discrepancies between manual and automatic estimates of RVEDA and RVESA. Interestingly, these errors do not propagate to RVFAC for this patient, as the overestimation of both areas partially cancels out. The resulting RVFAC estimation shows a bias and std of 0.19 ± 6.64 %, comparable to [17].

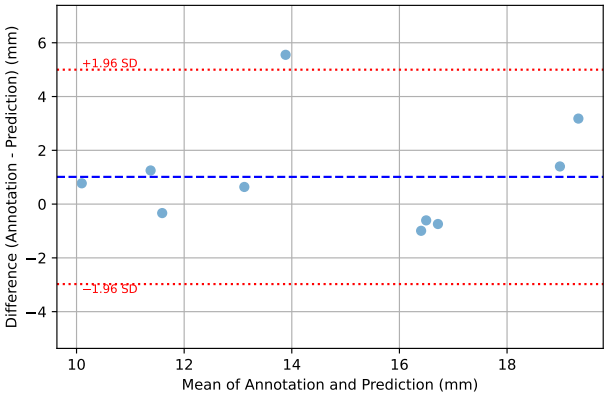
A similar behavior was observed for RVLSs. While absolute errors in RVLs were generally small relative to ventricular length, they compounded when computing strain. For example, although the normalized LoA for ED-RVL$_{fw}$ and ES-RVL$_{fw}$ were 22.5% and 34.3%, respectively, the normalized LoA for RVLS$_{fw}$ increased to 60.9%. This effect is illustrated by patient 149, in whom a slight displacement of TA^{ED}_{sep} leads to overestimation of septal, midpoint, and global longitudinal strains while RVLS$_{fw}$ remains unaffected (fig. 17).



(a) TAPSE$_{fw}$

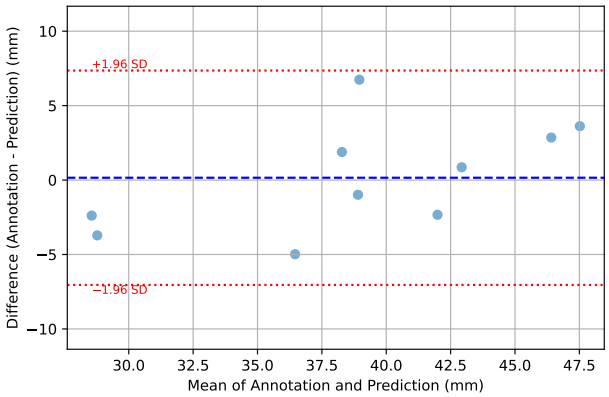


(b) TAPSE$_{sep}$

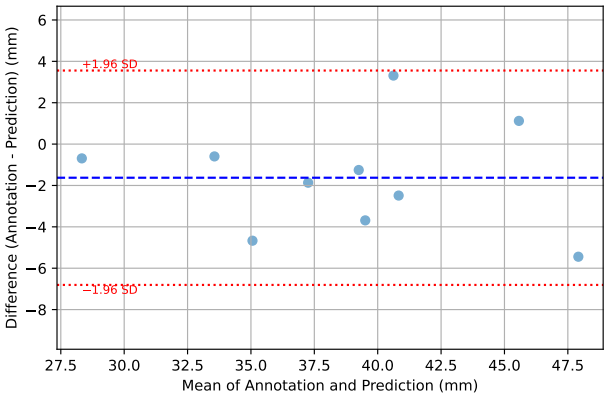


(c) TAPSE$_{global}$

Figure 13: Bland-Altman plots for TAPSE indices over the test set.

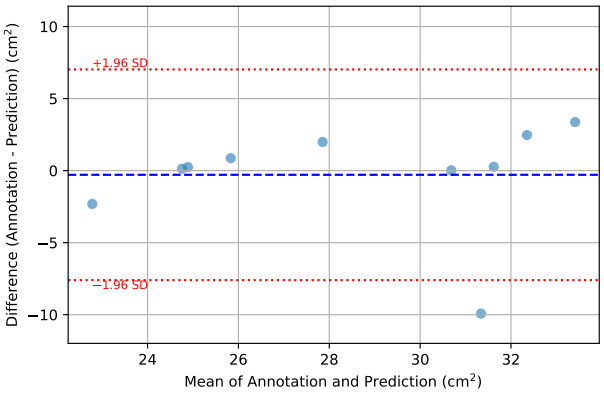


(a) TADD

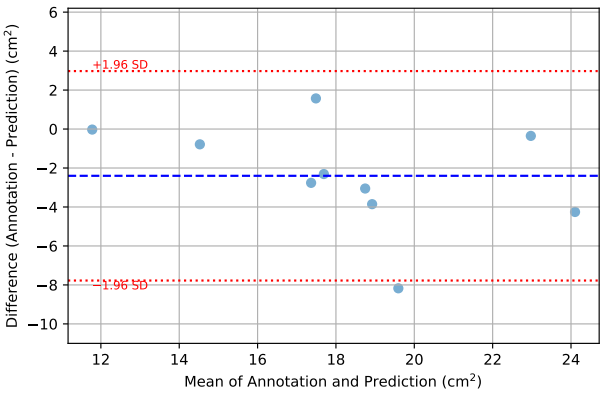


(b) TASD

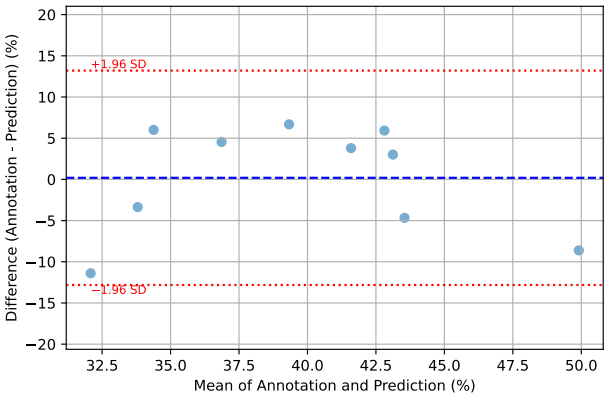
Figure 14: Bland-Altman plots for TAD indices over the test set.



(a) RVEDA

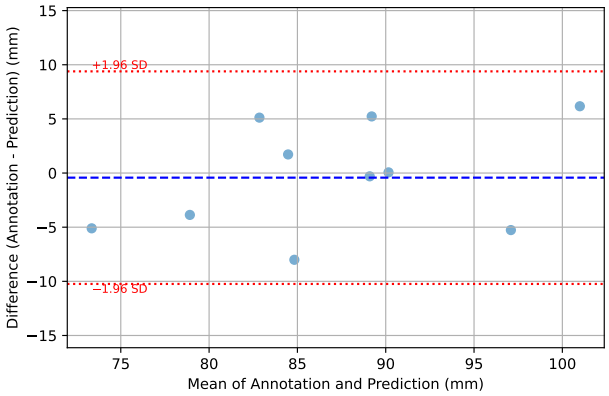


(b) RVESA

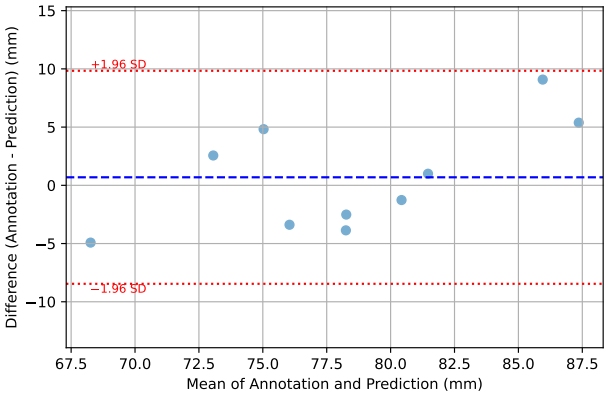


(c) RVFAC

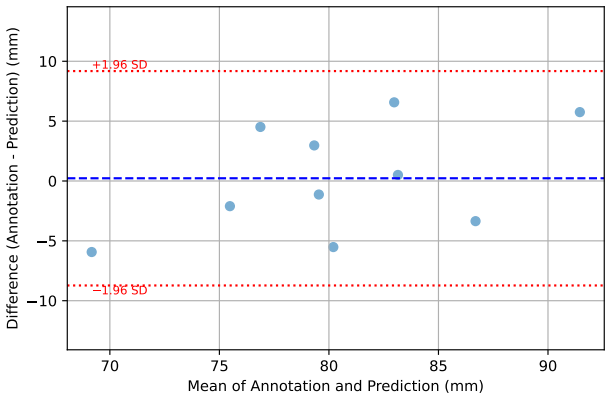
Figure 15: Bland-Altman plots for RV area indices over the test set.



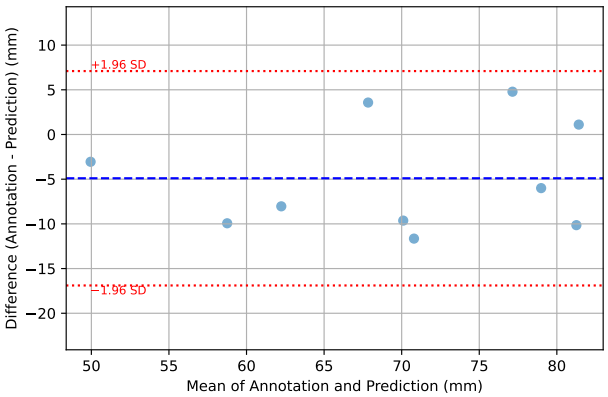
(a) ED-RVL$_{fw}$



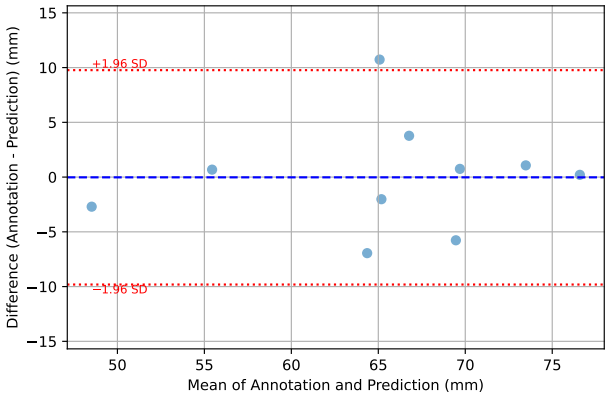
(b) ED-RVL$_{sep}$



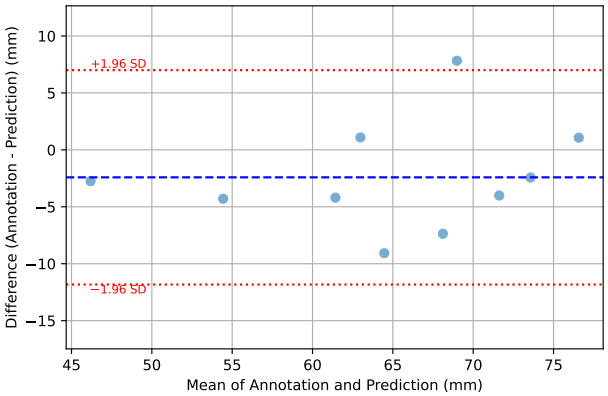
(c) ED-RVL$_{mid}$



(d) ES-RVL$_{fw}$

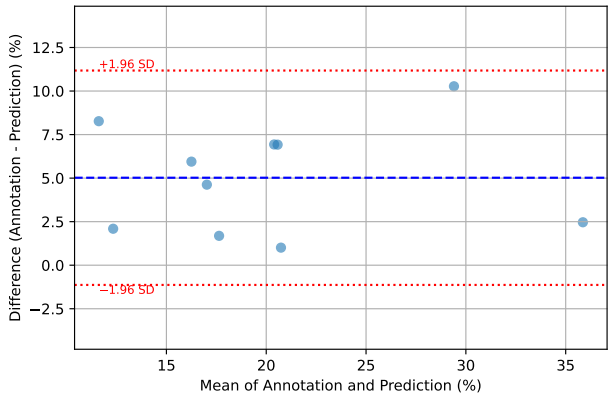


(e) ES-RVL$_{sep}$

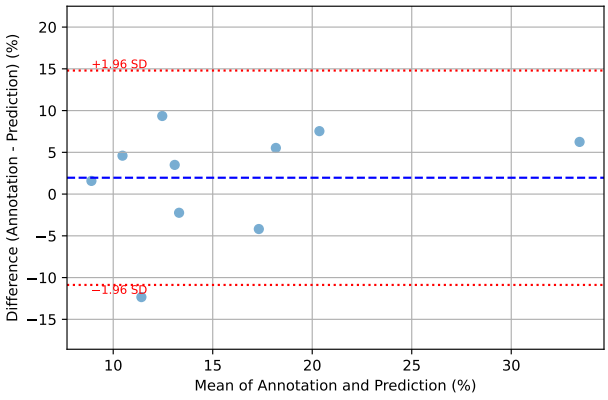


(f) ES-RVL$_{mid}$

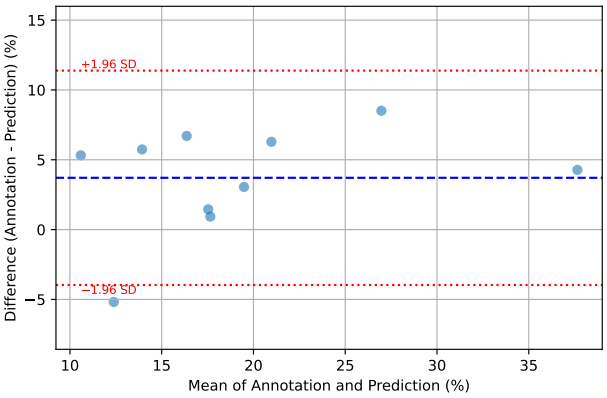
Figure 16: Bland-Altman plots for RV lengths over the test set.



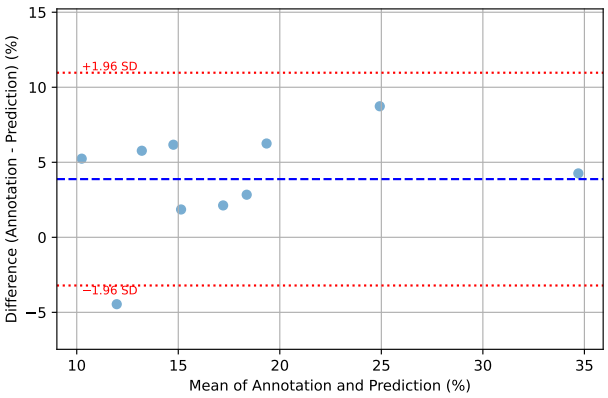
(a) RVLS_{fw}



(b) RVLS_{sep}



(c) RVLS_{mid}



(d) RVLS_{global}

Figure 17: Bland-Altman plots for RVLS indices over the test set.

Inference speed on test set

Network	Inference speed (*fps*)
U-Net	248
ResNet18	**287**
ResNet34	177
ResNet50	125
ResNeXt50	106
SwinUNETR	54

Table 8: Average inference speed measured on the test set. Best result shown in bold.

Clinical indices Mismatches

Group	Index family	Index	Value
TAPSE	TA motion	TAPSE_{fw}	1.09 ± 2.25 mm
		TAPSE_{sep}	0.93 ± 3.87 mm
		TAPSE_{global}	1.01 ± 2.03 mm
Diameters	TA distance	TADD	0.16 ± 3.67 mm
		TASD	-1.63 ± 2.64 mm
Areas	RV area	RVEDA	-0.29 ± 3.73 cm^2
		RVESA	2.40 ± 2.74 cm^2
		RVFAC	0.19 ± 6.64 %
DRVL	RV length	ED-RVL_{fw}	-0.42 ± 5.00 mm
		ED-RVL_{sep}	0.69 ± 4.66 mm
		ED-RVL_{mid}	0.22 ± 4.57 mm
SRVL	RV shortening	ES-RVL_{fw}	-4.89 ± 6.12 mm
		ES-RVL_{sep}	-0.02 ± 4.99 mm
		ES-RVL_{mid}	-2.41 ± 4.80 mm
RVS	Linear strain	RVLS_{fw}	5.02 ± 3.14 %
		RVLS_{sep}	1.96 ± 6.55 %
		RVLS_{mid}	3.71 ± 3.92 %
		RVLS_{global}	3.87 ± 3.62 %

Table 9: Mismatches (average ± standard deviation) between AutoRV-computed and manual clinical indices.

3.3. Threshold-based classification of RV function

To assess the ability of autoRV to detect RV dysfunction, the 10 RVs in the test set were classified based on TAPSE_{fw}, the best performing index among those reported in the literature. A conservative threshold of 18 mm [31] was adopted, at the upper end of the 15–18 mm range reported in previous studies [17, 20, 27, 32]. RV function was classified as impaired when TAPSE_{fw} was smaller than 18 mm. A relatively high threshold was intentionally selected to reduce the risk of missing severely impaired RV cases. Using a lower cutoff within the reported TAPSE_{fw} range could result in slight measurement errors causing truly pathological ventricles to be classified as normal.

Five RVs with normal function and four RVs with impaired function were correctly classified, while one normal RV was misclassified as impaired. No impaired RV was misclassified as normal (fig. 18).

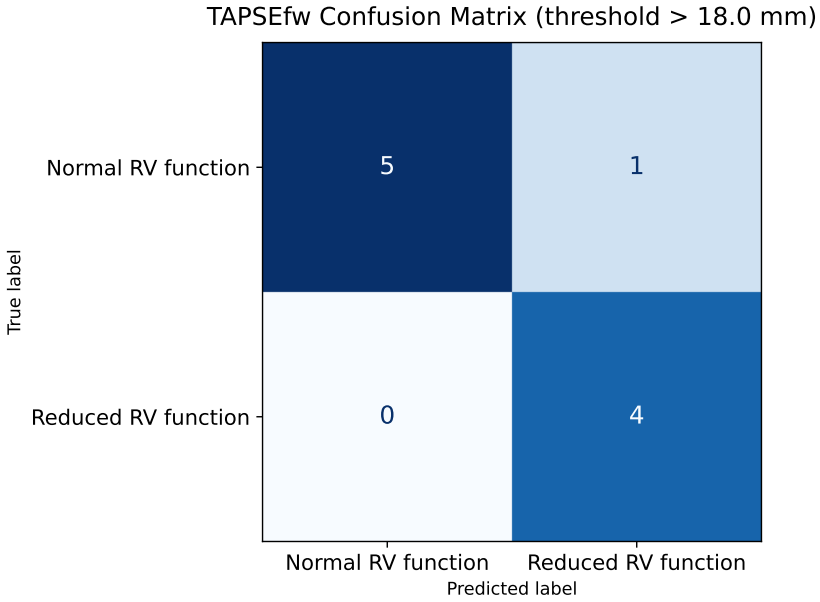


Figure 18: Confusion matrix for binary classification of RV function based on TAPSE_{fw} using a conservative threshold of 18 mm.

4. Discussion

The main finding of this study is that the deep learning pipeline herein presented, named AutoRV, automatically measures multiple RV parameters with very low bias. Data suggest that AutoRV shows promise for automatic monitoring of RV structure and function in mechanically ventilated patients.

4.1. AutoRV

The landmark detection task is well performed by our U-Net solution; however, for A, the distance between ground truth and predicted position can be as large as 30 mm. Additionally, TA_{fw} was localized more accurately than TA_{sep}. This highlights the difficulty of precisely and consistently localizing that landmark on 2D-TEE imaging. The larger error observed for A is mainly attributable to severe foreshortening and to substantial derangements in RV shape with respect to the subjects included in the training set. In contrast, the higher error for TA_{sep} is primarily related to foreshortening when the aortic valve is visible in the image (e.g., fig. 19), a condition that alters the apparent geometry of the annulus. Overall, more challenging cases were characterized by suboptimal image quality, which was reported in 20% of the acquisitions annotated by clinicians, by severe foreshortening, such as when TA_{fw} falls outside the image in certain time frames, and by marked RV shape alterations compared to the training population. Experiments were conducted, using filtering, to mitigate the high time-variance in the predictions. While filtering helps smooth out outliers, it has the side effect of erasing small subtleties in movements. Additionally, since consecutive frames are visually similar, any error in detecting a landmark usually propagates across frames, making filtering less effective.
Detection uncertainty naturally impacts the computation of clinical indixes. Although the proposed pipeline computes TAPSE_{fw} with narrower LoA than a recent semi-automated method [17], the observed errors remain substantially higher than expert inter- and intra-observer variability [76, 77]. This suggests that full automation is achieved at the cost of reduced agreement with manual measurements, highlighting a gap between current automated performance and expert-level consistency. The reduced stability of indices involving the septal annular point likely reflects challenges in accurately localizing TA_{sep}, whose errors propagate to derived indices and amplify variability compared to measurements based solely on TA_{fw}. Although TAPSE_{global} improves over TAPSE_{sep}, its performance remains similar to TAPSE_{fw}, while its mean value is lower, suggesting that combining annular landmarks does not substantially mitigate localization errors. The stable performance of TAD suggests robust geometric characterization of the annulus. This likely originates from the higher accuracy of the proposed method in detecting TA hinge points compared to A, as discussed in section 3.1. In contrast,

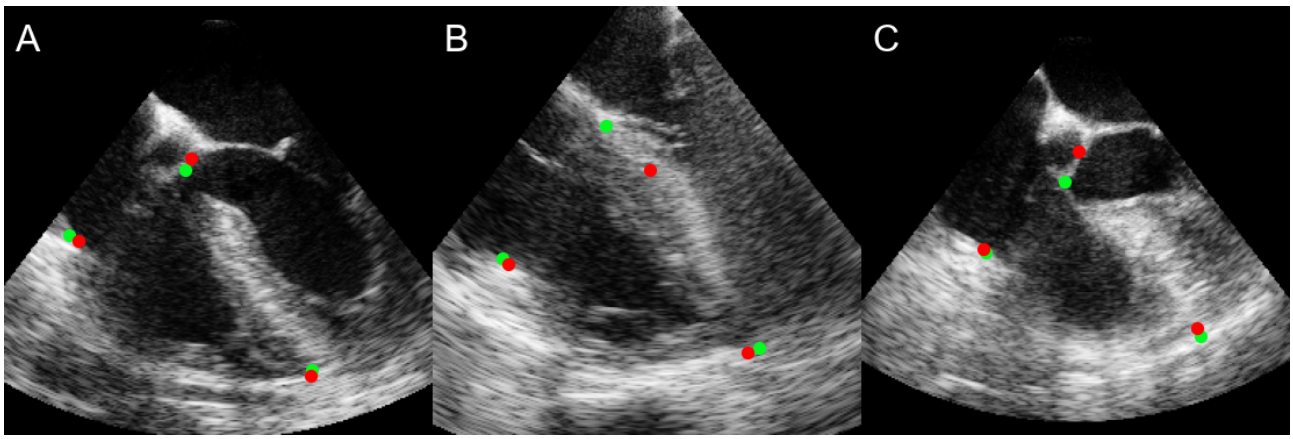


Figure 19: Three cases of poor image quality in the test set. The ground truth is reported in green, while the model predictions are the red dots. Images A and C, coming from patients 198 and 160 respectively, show how the presence of the aortic valve in the frame can create confusion in the model, that picks a point higher than the correct one. Image B, coming from patient 170, has a general lower quality, with less contrast than other acquisitions, so the model does not have enough patterns to find the correct landmark.

measurements relying on A, such as RVFAC, RVLs, and RVSs, are more susceptible to localization errors, leading to increased variability. One of the considered patients, i.e., patient ID: 190 (fig. 11), was shown to have a relatively narrow RV in comparison to the rest of our dataset, leading to clear outlier errors, with RVEDA and RVESA of 10.50 and 8.50 cm^2 respectively (fig. 15). A significant limitation lies in the approximation used to compute these areas: the cubic spline interpolation relies on the position of only three landmarks, implicitly assuming a standard curvature of the ventricular walls. Consequently, this method fails to capture local shape in case of morphological abnormalities. Although area errors may partially cancel in RVFAC, the metric remains sensitive to compounded errors, explaining the higher variability of this index. Similarly, strain metrics amplify small localization errors. Minor inaccuracies in detecting TA_{sep} and the apex propagate through longitudinal measurements and become magnified when computing relative deformation, explaining the large normalized LoA observed for strain indices.
These results highlight the lack of generalization across varying RV shapes and sizes from our model. A dataset with broader morphology representation is mandatory to achieve a more accurate RV function description. Indeed, though 5413 frames are used to train AutoRV, they depict only 36 different patients and thus morphologies.
The data used in this study were collected at a single hospital, which limits external validity and variations in image quality (e.g., foreshortening). The manual annotations were also performed by one clinician only, which prevents thorough comparisons between inter-observer variability and AutoRV.
Additionally, AutoRV relies on linear measurements in a single image plane and is therefore fundamentally limited by RV geometry. Moreover, our method does not account for inter- or intra-ventricular dyssynchrony, such as paradoxical septal motion, which may be important to detect during monitoring. Finally, although herein many indices are reported, the optimal combination of parameters for clinical use remains unclear.

4.2. Manual Measurements as a Baseline

When automating measurements on echocardiography, manually computed ones serve as reference ground truth. Therefore, mismatches between predictions by the model and manual measurements are interpreted as errors in the predictions. However, manual measurements are subject to their own uncertainty and variability: indeed, in this study landmarks were traced from scratch every time a linear length was manually measured. These factors probably widened the LoA regardless of the accuracy of AutoRV [78].

4.3. The Advantage of Multiple Parameters of the Right Ventricle

An advantage of the method herein presented is the automatic and real-time measurement of several parameters reflecting both RV function and size. Accounting for the latter helps interpreting the former, because changes in parameters of RV function may result from changes in preload, afterload, or contractility. By monitoring RV ED size, preload-mediated changes in RV function can be easily identified, and specific treatments can be

initiated [79].
Deciphering changes in RV function due to increased afterload or decreased contractility requires clinical context and cannot be achieved by interpreting TEE-derived measurements alone. A useful perspective is to view these load-dependent parameters of RV function as markers of right ventriculo-arterial coupling (i.e., the matching of contractility and afterload), which reflects the mechanical energy efficiency of the RV [80, 81]. As such, AutoRV may prove valuable for monitoring the evolution of RV function and the effects of therapeutic interventions in individual patients, when their hemodynamic state and contractility undergo significant variations over time.
There is also the question of whether to include septal measurements when assessing RV function, as our method does. Usually the septum is not considered when reporting RV function [40], because septal motion is influenced by LV contraction, which is a confounding factor in the analysis of RV function, whereas the RV free wall does not suffer from this limitation.
Whether analyzing the septum and free wall separately offers a clinical advantage likely depends on the clinical question. If one wishes to monitor a patient's global cardiopulmonary status, including the septum and LV function might be advantageous. On the other hand, if the specific question concerns cardiogenic versus non-cardiogenic pulmonary oedema, separating the left- and right-sided components could be useful [82]. However, whether such separated analysis makes sense when assessing global circulatory function is debatable: the septum is an integral part of normal RV function and accounts for a large proportion of global RV contractility [83].
Septal dysfunction is a key feature of RV dysfunction. Furthermore, RV dysfunction or dilatation can decrease LV preload and thereby reduce LV strain [84]. Thus, the RV and the LV are not only anatomically connected, but also functionally interdependent; attempts to separate their analysis may not be meaningful.

4.4. Threshold-Based Classification of RV Function

Using a conservative threshold of 18 mm to classify patients based on TAPSE_{fw}, i.e., the index most accurately predicted by AutoRV, showed strong classification performance. The results suggest that the method can reliably identify patients at risk of RV dysfunction, particularly by minimizing false negatives.
However, this result should be interpreted cautiously due to the limited cohort size and the fact that the analysis evaluates threshold-based agreement rather than true clinical diagnosis. Moreover, most established TAPSE_{fw} thresholds are derived from TTE, whereas the present measurements were obtained using TEE, where values tend to be slightly lower. This difference in imaging modality may influence classification performance.
Although binary classification shows promising accuracy, these results emphasize the need for continuous assessment of multiple indexes, rather than relying solely on a threshold-based method. Subtle declines in any of the indexes may may allow for earlier detection of RV dysfunction. Slight misclassification highlights the importance of a more precise estimation of the indices; even a few millimeters mismatch can result in incorrect decision making. Nonetheless, the findings support the prospective utility of AutoRV in identifying patients at risk of RV failure, emphasizing its future potential for guiding timely interventions and monitoring disease progression.

5. Conclusions and future developments

This thesis investigated the feasibility of a fully automated pipeline for real-time RV evaluation in mechanically ventilated patients using 2D TEE. To the best of our knowledge, this represents the first study to demonstrate an end-to-end system capable of automatically detecting anatomical landmarks, deriving clinically relevant indexes, and supporting continuous monitoring of RV function in this challenging clinical setting.
Several deep learning models were developed and compared for the automatic detection of the FW and septal TA hinge points and of the RV apex. Among them, a U-Net based architecture achieved the best performance, with mean localization errors on the order of 4.16 ± 3.36 mm on the test set. From the predicted landmark coordinates, the proposed autoRV pipeline automatically computed multiple quantitative indexes commonly used in RV assessment, including different variants of TAPSE, RV areas and RVFAC, RVLs, and RVLSs. Together, these parameters capture complementary aspects of RV mechanics and geometry, enabling a more comprehensive evaluation of RV function than any single index alone.
With respect to traditional STE, which remains the reference technique for RV strain analysis, the proposed approach offers several advantages. STE requires expert operator input during both acquisition and post-processing, is highly sensitive to image quality, and lacks standardization in TEE, particularly in mechanically ventilated patients. Moreover, most available implementations are limited to single-beat analysis, reducing robustness and limiting their usefulness for monitoring temporal changes. In contrast, the autoRV pipeline operates fully automatically, supports continuous beat-by-beat analysis, and naturally enables averaging over multiple cardiac cycles. This capability improves measurement robustness and enhances the interpretability of longitudinal changes in RV performance, which is crucial in critical care environments.

Despite these promising results, several limitations remain. High image quality is still a prerequisite, as artifacts, foreshortening, and the appearance of the aortic valve can degrade landmark detection and index accuracy. Although the proposed method shows good agreement with clinical measurements for several parameters, its accuracy is not yet sufficient to reliably detect subtle impairments in RV function, and performance remains inferior to expert assessment. This discrepancy indicates that the current level of error is still too large to be considered clinically reliable for independent assessment of RV function without expert supervision. Indices relying on TA_{sep} and A remain more sensitive to localization errors and currently show limited robustness compared to measurements based on TA_{fw}. While annular geometry can be estimated reliably, composite functional indices such as RVFAC and RVLS remain challenging due to error propagation and anatomical variability. Moreover, a limitation is represented by how the RV areas are computed and their implicit assumption of a standard curvature of the RV walls. In addition, most of the considered indexes are standardized and validated primarily in TTE, whereas their interpretation and threshold values in TEE are less well established. As a consequence, a robust and clinically validated strategy for aggregating multiple predicted indexes into a unified assessment of RV performance is still missing.

Future work should therefore focus on improving the stability of landmark detection and the accuracy of landmark frame-to-frame tracking, for instance by incorporating temporal information through architectures that jointly process short sequences of frames. A promising direction would be to train the model to output landmark predictions only when its confidence is sufficiently high. This would improve overall accuracy, reduce sensitivity to image quality, and limit the occurrence of gross errors. Accuracy and precision could also be enhanced by systematically averaging measurements over a larger number of cardiac cycles, ideally ten or more, to facilitate earlier detection of meaningful physiological changes. Implementing the segmentation of the RV chamber, whose isosurface would correspond to the RV endocardium ([17]), could also be a great improvement. Traditional segmentation-based approaches are likely to provide more robust geometric estimates of RV area and make apex detection easier, although at the cost of substantially increased annotation effort compared to the present keypoint-based strategy. Furthermore, the development of a dedicated AI-based classifier that integrates all predicted indexes could enable automatic detection of subtle functional variations and provide early alerts to clinicians, by learning to evaluate RV function based on the optimal combination of clinical indices. Extensions toward 3D TEE based models could further improve geometric fidelity by capturing the full spatial complexity of RV anatomy, though this would increase model complexity, processing time and computational burden. Future validation studies could also address the inherent variability of manual measurements. This issue may be addressed in three ways. First, a third, independent "umpire test" could be used as a term of comparison for both AutoRV and manual measurements ([85]). Second, a repeated-measurement study could be performed to assess the test–retest variability of both methods, thereby evaluating a more relevant characteristic for monitoring in settings with an imperfect reference ([86]). Third, tracings and measurements could be averaged over a larger number of cardiac cycles, thus mitigating the impact of spurious errors and potentially making the LoA narrower ([66, 67]). These directions should be explored in a future validation study. Finally, training and validation on larger and more diverse patient cohorts will be essential to ensure generalization, statistical robustness, and clinical relevance.

In conclusion, this work demonstrates that automated monitoring of RV function in 2D-TEE is technically feasible in mechanically ventilated patients. By leveraging a lightweight U-Net architecture for automatic landmark detection and deriving multiple clinical indexes in real time, the proposed autoRV pipeline offers a standardized, objective, and operator-independent framework for RV assessment. Its fast inference time and minimal computational requirements make it realistically applicable in the ICU. While further optimization and large-scale validation are needed, this study lays the groundwork for a new generation of echocardiographic tools capable of continuous and reproducible cardiac monitoring in demanding critical care settings.

References

[1] Elio Aloia, Matteo Cameli, Flavio D'Ascenzi, Carlotta Sciaccaluga, and Sergio Mondillo. TAPSE: An old but useful tool in different diseases. *International Journal of Cardiology*, 225:177–183, December 2016.

[2] Samantha K. Brenner, Tariq U. Azam, Joseph E. Parrillo, Steven M. Hollenberg, Elizabeth Anderson, Patrick O'Hayer, Hanna Berlin, Pennelope Blakley, Abbas Bitar, and Salim S. Hayek. Right Ventricular Dysfunction in Critically Ill Patients With COVID-19. *The American Journal of Cardiology*, 168:176–178, April 2022.

[3] Michael J. Lanspa, Meghan M. Cirulis, Brandon M. Wiley, Troy D. Olsen, Emily L. Wilson, Sarah J. Beesley, Samuel M. Brown, Eliotte L. Hirshberg, and Colin K. Grissom. Right Ventricular Dysfunction in Early Sepsis and Septic Shock. *Chest*, 159(3):1055–1063, March 2021.

[4] Hongmin Zhang, Wei Huang, Qing Zhang, Xiukai Chen, Xiaoting Wang, Dawei Liu, and Critical Care

Ultrasound Study Group. Prevalence and prognostic value of various types of right ventricular dysfunction in mechanically ventilated septic patients. *Ann. Intensive Care*, 11(1):108, December 2021.

[5] Matthieu Petit, Edouard Jullien, and Antoine Vieillard-Baron. Right Ventricular Function in Acute Respiratory Distress Syndrome: Impact on Outcome, Respiratory Strategy and Use of Veno-Venous Extracorporeal Membrane Oxygenation. *Front Physiol*, 12:797252, January 2022.

[6] Ryota Sato, Siddharth Dugar, Wisit Cheungpasitporn, Mary Schleicher, Patrick Collier, Saraschandra Vallabhajosyula, and Abhijit Duggal. The impact of right ventricular injury on the mortality in patients with acute respiratory distress syndrome: a systematic review and meta-analysis. *Crit Care*, 25(1):172, December 2021.

[7] Antoine Vieillard-Baron, R. Naeije, F. Haddad, H. J. Bogaard, T. M. Bull, N. Fletcher, T. Lahm, S. Magder, S. Orde, G. Schmidt, and M. R. Pinsky. Diagnostic workup, etiologies and management of acute right ventricle failure: A state-of-the-art paper. *Intensive Care Med*, 44(6):774–790, June 2018.

[8] Simone Cappio Borlino, Julien Hagry, Christopher Lai, Eduardo Rocca, Gaëlle Fouqué, Daniela Rosalba, Marta Fasan, Rui Shi, Agnese Recanatini, Irene Cisterna, Mattia Barotti, Tài Pham, Jean-Louis Teboul, and Xavier Monnet. The Effect of Positive End-Expiratory Pressure on Pulmonary Vascular Resistance Depends on Lung Recruitability in Patients with Acute Respiratory Distress Syndrome. *Am J Respir Crit Care Med*, 210(7):900–907, October 2024.

[9] Armand Mekontso Dessap, Florence Boissier, Cyril Charron, Emmanuelle Bégot, Xavier Repessé, Annick Legras, Christian Brun-Buisson, Philippe Vignon, and Antoine Vieillard-Baron. Acute cor pulmonale during protective ventilation for acute respiratory distress syndrome: prevalence, predictors, and clinical impact. *Intensive Care Med*, 42(5):862–870, May 2016.

[10] Antoine Vieillard-Baron, Amélie Prigent, Xavier Repessé, Marine Goudelin, Gwenaël Prat, Bruno Evrard, Cyril Charron, Philippe Vignon, and Guillaume Geri. Right ventricular failure in septic shock: characterization, incidence and impact on fluid responsiveness. *Crit Care*, 24(1):630, December 2020.

[11] Juan C. Grignola and Enric Domingo. Acute Right Ventricular Dysfunction in Intensive Care Unit. *BioMed Research International*, 2017:1–15, 2017.

[12] Laura C. Price, Guillermo Martinez, Aimee Brame, Thomas Pickworth, Chinthaka Samaranayake, David Alexander, Benjamin Garfield, Tuan-Chen Aw, Colm McCabe, Bhashkar Mukherjee, Carl Harries, Aleksander Kempny, Michael Gatzoulis, Philip Marino, David G. Kiely, Robin Condliffe, Luke Howard, Rachel Davies, Gerry Coghlan, Benjamin E. Schreiber, James Lordan, Dolores Taboada, Sean Gaine, Martin Johnson, Colin Church, Samuel V. Kemp, Davina Wong, Andrew Curry, Denny Levett, Susanna Price, Stephane Ledot, Anna Reed, Konstantinos Dimopoulos, and Stephen John Wort. Perioperative management of patients with pulmonary hypertension undergoing non-cardiothoracic, non-obstetric surgery: a systematic review and expert consensus statement. *British Journal of Anaesthesia*, 126(4):774–790, April 2021.

[13] Nawaporn Assanangkornchai, Valerie Villeneuve, Sarah McDonald, Dominique Shum Tim, Sheldon Magder, Pattra Mettasittigorn, and Roupen Hatzakorzian. Reduction in Postoperative Right Ventricular Echocardiographic Indices Predicts Longer Duration of Vasoactive Support After Cardiac Surgery. *Journal of Cardiothoracic and Vascular Anesthesia*, 39(1):143–150, January 2025.

[14] Ramesh M. Gowda, Ijaz A. Khan, Balendu C. Vasavada, Terrence J. Sacchi, and Riti Patel. History of the evolution of echocardiography. *International Journal of Cardiology*, 97(1):1–6, October 2004.

[15] Caroline Bleakley and Mark Monaghan. 3D transesophageal echocardiography in TAVR. *Echocardiography*, 37(10):1654–1664, 2020. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/echo.14777.

[16] Edwin Wilberforce Turton and Jörg Ender. Role of 3D Echocardiography in Cardiac Surgery: Strengths and Limitations. *Curr Anesthesiol Rep*, 7(3):291–298, 2017.

[17] Artem Chernyshov, Jahn Frederik Grue, John Nyberg, Bjørnar Grenne, Håvard Dalen, Svein Arne Aase, Andreas Østvik, and Lasse Lovstakken. Automated Segmentation and Quantification of the Right Ventricle in 2-D Echocardiography. *Ultrasound in Medicine & Biology*, 50(4):540–548, April 2024.

[18] Davide Genovese, Nina Rashedi, Lynn Weinert, Akhil Narang, Karima Addetia, Amit R. Patel, David Prater, Alexandra Gonçalves, Victor Mor-Avi, and Roberto M. Lang. Machine Learning–Based Three-Dimensional Echocardiographic Quantification of Right Ventricular Size and Function: Validation Against Cardiac Magnetic Resonance. *Journal of the American Society of Echocardiography*, 32(8):969–977, August 2019.

[19] Rebecca T. Hahn, Stamatios Lerakis, Victoria Delgado, Karima Addetia, Daniel Burkhoff, Denisa Muraru, Sean Pinney, and Mark K. Friedberg. Multimodality Imaging of Right Heart Function. *Journal of the American College of Cardiology*, 81(19):1954–1973, May 2023.

[20] Lawrence G. Rudski, Wyman W. Lai, Jonathan Afilalo, Lanqi Hua, Mark D. Handschumacher, Krishnaswamy Chandrasekaran, Scott D. Solomon, Eric K. Louie, and Nelson B. Schiller. Guidelines for the Echocardiographic Assessment of the Right Heart in Adults: A Report from the American Society of Echocardiography. *Journal of the American Society of Echocardiography*, 23(7):685–713, July 2010.

[21] Carole A. Warnes. Adult Congenital Heart Disease. *Journal of the American College of Cardiology*, 54(21):1903–1910, November 2009.

[22] Lee Fong Ling, Nancy A. Obuchowski, Leonardo Rodriguez, Zoran Popovic, Deborah Kwon, and Thomas H. Marwick. Accuracy and Interobserver Concordance of Echocardiographic Assessment of Right Ventricular Size and Systolic Function: A Quality Control Exercise. *Journal of the American Society of Echocardiography*, 25(7):709–713, July 2012.

[23] Sam Orde, Michel Slama, Konstantin Yastrebov, Anthony Mclean, and Stephen Huang. Subjective right ventricle assessment by echo qualified intensive care specialists: assessing agreement with objective measures. *Crit Care*, 23:70, March 2019.

[24] Sanjiv Kaul, Chuwa Tei, James M. Hopkins, and Pravin M. Shah. Assessment of right ventricular function using two-dimensional echocardiography. *American Heart Journal*, 107(3):526–531, March 1984.

[25] George T. Karatasakis, Labros A. Karagounis, Periklis A. Kalyvas, Athanassios Manginas, George D. Athanassopoulos, Stefanos A. Aggelakas, and Dennis V. Cokkinos. Prognostic significance of echocardiographically estimated right ventricular shortening in advanced heart failure. *American Journal of Cardiology*, 82(3):329–334, August 1998. Publisher: Elsevier.

[26] Paul R. Forfia, Micah R. Fisher, Stephen C. Mathai, Traci Housten-Harris, Anna R. Hemnes, Barry A. Borlaug, Elzbieta Chamera, Mary C. Corretti, Hunter C. Champion, Theodore P. Abraham, Reda E. Girgis, and Paul M. Hassoun. Tricuspid Annular Displacement Predicts Survival in Pulmonary Hypertension. *Am J Respir Crit Care Med*, 174(9):1034–1041, November 2006. Publisher: American Thoracic Society - AJRCCM.

[27] Stefano Ghio, Catherine Klersy, Giulia Magrini, Andrea Maria D'Armini, Laura Scelsi, Claudia Raineri, Michele Pasotti, Alessandra Serio, Carlo Campana, and Mario Viganò. Prognostic relevance of the echocardiographic assessment of right ventricular function in patients with idiopathic pulmonary arterial hypertension. *International Journal of Cardiology*, 140(3):272–278, April 2010.

[28] L. P. Badano, C. Ginghina, J. Easaw, D. Muraru, M. T. Grillo, P. Lancellotti, B. Pinamonti, G. Coghlan, M. P. Marra, B. A. Popescu, and S. De Vita. Right ventricle in pulmonary arterial hypertension: haemodynamics, structural changes, imaging, and proposal of a study protocol aimed to assess remodelling and treatment effects. *European Journal of Echocardiography*, 11(1):27–37, January 2010.

[29] Shayne Michael Roberts, John Klick, Adrian Fischl, Tonya S. King, and Theodore J. Cios. A Comparison of Transesophageal to Transthoracic Echocardiographic Measures of Right Ventricular Function. *Journal of Cardiothoracic and Vascular Anesthesia*, 34(5):1252–1259, May 2020.

[30] Davide Genovese, Victor Mor-Avi, Chiara Palermo, Denisa Muraru, Valentina Volpato, Eric Kruse, Megan Yamat, Patrizia Aruta, Karima Addetia, Luigi P. Badano, and Roberto M. Lang. Comparison Between Four-Chamber and Right Ventricular–Focused Views for the Quantitative Evaluation of Right Ventricular Size and Function. *Journal of the American Society of Echocardiography*, 32(4):484–494, April 2019.

[31] Eckhard Schmid, Jan N Hilberath, Gunnar Blumenstock, Prem S Shekar, Steffen Kling, Stanton K Shernan, Peter Rosenberger, and Martina Nowak-Machen. Tricuspid annular plane systolic excursion (TAPSE) predicts poor outcome in patients undergoing acute pulmonary embolectomy. *Heart Lung Vessel*, 7(2):151–158, 2015.

[32] Shadi Lahham, John C Fox, Maxwell Thompson, Tanyaporn Nakornchai, Badriah Alruwaili, Ghadeer Doman, Shannon May Lee, Amal Shafi, Inna Shniter, Victoria Valdes, and Lishi Zhang. Tricuspid Annular Plane of Systolic Excursion to Prognosticate Acute Pulmonary Symptomatic Embolism (TAPSEPAPSE Study). *J Ultrasound Med*, 38(3):695–702, March 2019.

[33] Xiang Ji, Yiwei Zhang, Yuji Xie, Wenyuan Wang, Yuting Tan, Mingxing Xie, and Li Zhang. Feasibility Value of Right Ventricular Longitudinal Shortening Fraction and the Prognostic Implications in Patients With Heart Transplantation. *JAHA*, 13(6):e032402, March 2024.

[34] Denisa Muraru, Kristina Haugaa, Erwan Donal, Ivan Stankovic, Jens Uwe Voigt, Steffen E Petersen, Bogdan A Popescu, and Thomas Marwick. Right ventricular longitudinal strain in the clinical routine: a state-of-the-art review. *European Heart Journal - Cardiovascular Imaging*, 23(7):898–912, June 2022.

[35] Denisa Muraru, Sebastian Onciul, Diletta Peluso, Nicola Soriani, Umberto Cucchini, Patrizia Aruta, Gabriella Romeo, Giacomo Cavalli, Sabino Iliceto, and Luigi P. Badano. Sex- and Method-Specific Reference Values for Right Ventricular Strain by 2-Dimensional Speckle-Tracking Echocardiography. *Circ: Cardiovascular Imaging*, 9(2), February 2016.

[36] Philippe Unger, Marianne Paesmans, Jean-Luc Vachiery, Michael Rietz, Mihaela Amzulescu, and Aurelia David-Cojocariu. Right ventricular longitudinal fractional shortening: a substitute to right ventricular free wall longitudinal strain? *Heart Vessels*, 37(3):426–433, March 2022.

[37] Homaa Ahmad, Victor Mor-Avi, Roberto M. Lang, Hans-Joachim Nesser, Lynn Weinert, Wendy Tsang, Regina Steringer-Mascherbauer, Johannes Niel, Ivan S. Salgo, and Lissa Sugeng. Assessment of Right Ventricular Function Using Echocardiographic Speckle Tracking of the Tricuspid Annular Motion: Comparison with Cardiac Magnetic Resonance. *Echocardiography*, 29(1):19–24, 2012. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-8175.2011.01519.x.

[38] Weiting Huang, James Hodovan, Avneesh Sharma, Matteo Morello, Onur Varli, Bethany Gholson, and Jonathan R. Lindner. Impact of Length Indexing of Deformation in Echocardiographic Evaluation of Right Ventricular Function. *Journal of the American Society of Echocardiography*, 38(3):187–194, March 2025.

[39] Julien Dreyfus, Guillaume Durand-Viel, Richard Raffoul, Soleiman Alkhoder, Ulrik Hvass, Costin Radu, Nawwar Al-Attar, Walid Ghodbhane, David Attias, Patrick Nataf, Alec Vahanian, and David Messika-Zeitoun. Comparison of 2-Dimensional, 3-Dimensional, and Surgical Measurements of the Tricuspid Annulus Size: Clinical Implications. *Circ: Cardiovascular Imaging*, 8(7):e003241, July 2015.

[40] Monica Mukherjee, Lawrence G. Rudski, Karima Addetia, Jonathan Afilalo, Michele D'Alto, Benjamin H. Freed, Lynsy B. Friend, Luna Gargani, Julia Grapsa, Paul M. Hassoun, Lanqi Hua, Jiwon Kim, Valentina Mercurio, Rajan Saggar, and Anton Vonk-Noordegraaf. Guidelines for the Echocardiographic Assessment of the Right Heart in Adults and Special Considerations in Pulmonary Hypertension: Recommendations from the American Society of Echocardiography. *Journal of the American Society of Echocardiography*, 38(3):141–186, March 2025.

[41] Anders Austlid Taskén, Erik Andreas Rye Berg, Bjørnar Grenne, Espen Holte, Håvard Dalen, Stian Stølen, Frank Lindseth, Svend Aakhus, and Gabriel Kiss. Automated estimation of mitral annular plane systolic excursion by artificial intelligence from 3D ultrasound recordings. *Artificial Intelligence in Medicine*, 144:102646, October 2023.

[42] Matteo Cameli, Giulia E. Mandoli, Carlotta Sciaccaluga, and Sergio Mondillo. More than 10 years of speckle tracking echocardiography: Still a novel technique or a definite tool for clinical practice? *Echocardiography*, 36(5):958–970, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/echo.14339.

[43] Sergio Mondillo, Maurizio Galderisi, Donato Mele, Matteo Cameli, Vincenzo Schiano Lomoriello, Valerio Zacà, Piercarlo Ballo, Antonello D'Andrea, Denisa Muraru, Mariangela Losi, Eustachio Agricola, Arcangelo D'Errico, Simona Buralli, Susanna Sciomer, Stefano Nistri, and Luigi Badano. Speckle-Tracking Echocardiography. *Journal of Ultrasound in Medicine*, 30(1):71–83, 2011. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.7863/jum.2011.30.1.71.

[44] Dorota Smolarek, Wojciech Sobiczewski, Maria Dudziak, and Marcin Hellmann. Speckle-tracking echocardiographic evaluation of the right ventricle in patients with ischemic left ventricular dysfunction. *Cardiol J*, 30(1):73–81, February 2023.

[45] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[46] Pallabi Sharma, Deepak Ranjan Nayak, Bunil Kumar Balabantaray, M. Tanveer, and Rajashree Nayak. A survey on cancer detection via convolutional neural networks: Current challenges and future directions. *Neural Networks*, 169:637–659, January 2024.

[47] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition, December 2015. arXiv:1512.03385 [cs].

[48] Karen Simonyan and Andrew Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition, April 2015. arXiv:1409.1556 [cs].

[49] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated Residual Transformations for Deep Neural Networks, April 2017. arXiv:1611.05431 [cs].

[50] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going Deeper with Convolutions, September 2014. arXiv:1409.4842 [cs].

[51] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, volume 9351, pages 234–241. Springer International Publishing, Cham, 2015. Series Title: Lecture Notes in Computer Science.

[52] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention Is All You Need, August 2023. arXiv:1706.03762 [cs].

[53] Niki Parmar, Ashish Vaswani, Jakob Uszkoreit, Łukasz Kaiser, Noam Shazeer, Alexander Ku, and Dustin Tran. Image Transformer, June 2018. arXiv:1802.05751 [cs].

[54] Rewon Child, Scott Gray, Alec Radford, and Ilya Sutskever. Generating Long Sequences with Sparse Transformers, April 2019. arXiv:1904.10509 [cs].

[55] Dirk Weissenborn, Oscar Täckström, and Jakob Uszkoreit. Scaling Autoregressive Video Models, February 2020. arXiv:1906.02634 [cs].

[56] Irwan Bello, Barret Zoph, Ashish Vaswani, Jonathon Shlens, and Quoc V. Le. Attention Augmented Convolutional Networks, September 2020. arXiv:1904.09925 [cs].

[57] Mark Chen, Alec Radford, Rewon Child, Jeff Wu, Heewoo Jun, Prafulla Dhariwal, David Luan, and Ilya Sutskever. Generative Pretraining from Pixels.

[58] Jean-Baptiste Cordonnier, Andreas Loukas, and Martin Jaggi. On the Relationship between Self-Attention and Convolutional Layers, January 2020. arXiv:1911.03584 [cs].

[59] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale, June 2021. arXiv:2010.11929 [cs].

[60] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, August 2021. arXiv:2103.14030 [cs].

[61] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection, April 2017. arXiv:1612.03144 [cs].

[62] Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger Roth, and Daguang Xu. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images, January 2022. arXiv:2201.01266 [eess].

[63] Erik Andreas Rye Berg, Anders Austlid Taskén, Trym Nordal, Bjørnar Grenne, Torvald Espeland, Idar Kirkeby-Garstad, Håvard Dalen, Espen Holte, Stian Stølen, Svend Aakhus, and Gabriel Kiss. Fully automatic estimation of global left ventricular systolic function using deep learning in transoesophageal echocardiography. *European Heart Journal - Imaging Methods and Practice*, 1(1):qyad007, May 2023.

[64] Anders Austlid Taskén, Jinyang Yu, Erik Andreas Rye Berg, Bjørnar Grenne, Espen Holte, Håvard Dalen, Stian Stølen, Frank Lindseth, Svend Aakhus, and Gabriel Kiss. Automatic Detection and Tracking of Anatomical Landmarks in Transesophageal Echocardiography for Quantification of Left Ventricular Function. *Ultrasound in Medicine & Biology*, 50(6):797–804, June 2024.

[65] Jinyang Yu, Anders Austlid Taskén, Erik Andreas Rye Berg, Tomas Dybos Tannvik, Katrine Hordnes Slagsvold, Idar Kirkeby-Garstad, Bjørnar Grenne, Gabriel Kiss, and Svend Aakhus. Continuous monitoring of left ventricular function in postoperative intensive care patients using artificial intelligence and transesophageal echocardiography. *ICMx*, 12(1):54, June 2024.

[66] Jinyang Yu, Anders Austlid Taskén, Erik Andreas Rye Berg, Tomas Dybos Tannvik, Katrine Hordnes Slagsvold, Idar Kirkeby–Garstad, Bjørnar Grenne, Gabriel Kiss, and Svend Aakhus. Continuous three-dimensional transesophageal echocardiography and deep learning for perioperative monitoring of left ventricular longitudinal function. *European Heart Journal - Imaging Methods and Practice*, 3(1):qyaf052, January 2025.

[67] Jinyang Yu, Anders Austlid Taskén, Hans Martin Flade, Eirik Skogvoll, Erik Andreas Rye Berg, Bjørnar Grenne, Audun Rimehaug, Idar Kirkeby-Garstad, Gabriel Kiss, and Svend Aakhus. Automatic assessment of left ventricular function for hemodynamic monitoring using artificial intelligence and transesophageal echocardiography. *J Clin Monit Comput*, 38(2):281–291, April 2024.

[68] Ashley N. Beecy, Alex Bratt, Brian Yum, Razia Sultana, Mukund Das, Ines Sherifi, Richard B. Devereux, Jonathan W. Weinsaft, and Jiwon Kim. Development of novel machine learning model for right ventricular quantification on echocardiography—A multimodality validation study. *Echocardiography*, 37(5):688–697, May 2020.

[69] Rohan Shad, Nicolas Quach, Robyn Fong, Patpilai Kasinpila, Cayley Bowles, Miguel Castro, Ashrith Guha, Erik E. Suarez, Stefan Jovinge, Sangjin Lee, Theodore Boeve, Myriam Amsallem, Xiu Tang, Francois Haddad, Yasuhiro Shudo, Y. Joseph Woo, Jeffrey Teuteberg, John P. Cunningham, Curtis P. Langlotz, and William Hiesinger. Predicting post-operative right ventricular failure using video-based deep learning. *Nat Commun*, 12(1):5192, August 2021.

[70] Márton Tokodi, Bálint Magyar, András Soós, Masaaki Takeuchi, Máté Tolvaj, Bálint Károly Lakatos, Tetsuji Kitano, Yosuke Nabeshima, Alexandra Fábián, Mark Bence Szigeti, András Horváth, Béla Merkely, and Attila Kovács. Deep Learning-Based Prediction of Right Ventricular Ejection Fraction Using 2D Echocardiograms. *JACC: Cardiovascular Imaging*, 16(8):1005–1018, August 2023.

[71] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices, December 2017. arXiv:1707.01083 [cs].

[72] Polydoros N. Kampaktsis, Tuan A. Bohoran, Mark Lebehn, Laura McLaughlin, Jay Leb, Zhonghua Liu, Serafeim Moustakidis, Athanasios Siouras, Anvesha Singh, Rebecca T. Hahn, Gerry P. McCann, and Archontis Giannakidis. An attention-based deep learning method for right ventricular quantification using 2D echocardiography: Feasibility and accuracy. *Echocardiography*, 41(1):e15719, January 2024.

[73] Jinyang Yu, Kjell Høyland, Erik Andreas Rye Berg, Tomas Dybos Tannvik, Bjørnar Grenne, and Svend Aakhus. Right Ventricular Ejection Fraction Assessed by Single-Beat Three-Dimensional Transesophageal Echocardiography in Mechanically Ventilated Patients. *Anesthesia & Analgesia*, March 2025.

[74] M. Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, Vishwesh Nath, Yufan He, Ziyue Xu, Ali Hatamizadeh, Andriy Myronenko, Wentao Zhu, Yun Liu, Mingxin Zheng, Yucheng Tang, Isaac Yang, Michael Zephyr, Behrooz Hashemian, Sachidanand Alle, Mohammad Zalbagi Darestani, Charlie Budd, Marc Modat, Tom Vercauteren, Guotai Wang, Yiwen Li, Yipeng Hu, Yunguan Fu, Benjamin Gorman, Hans Johnson, Brad Genereaux, Barbaros S. Erdal, Vikash Gupta, Andres Diaz-Pinto, Andre Dourson, Lena Maier-Hein, Paul F. Jaeger, Michael Baumgartner, Jayashree Kalpathy-Cramer, Mona Flores, Justin Kirby, Lee A. D. Cooper, Holger R. Roth, Daguang Xu, David Bericat, Ralf Floca, S. Kevin Zhou, Haris Shuaib, Keyvan Farahani, Klaus H. Maier-Hein, Stephen Aylward, Prerna Dogra, Sebastien Ourselin, and Andrew Feng. MONAI: An open-source framework for deep learning in healthcare, November 2022. arXiv:2211.02701 [cs].

[75] Jiapu Pan and Willis J. Tompkins. A Real-Time QRS Detection Algorithm. *IEEE Trans. Biomed. Eng.*, BME-32(3):230–236, March 1985.

[76] Mamatha Munaf, Puthuvassery Raman Suneel, Sivadasanpillai Harikrishnan, Deepa Sasikumar, and Thomas Koshy. Tricuspid Annular Plane Systolic Excursion (TAPSE) for the Assessment of Right Ventricular Function in Adult and Pediatric Cardiac Surgery: Modified Two-dimensional and M-mode TAPSE by Transesophageal Echocardiography Compared to M-mode TAPSE by Transthoracic Echocardiography. *Journal of Cardiothoracic and Vascular Anesthesia*, 38(1):123–132, January 2024.

[77] Henry Skinner, Hazlyna Kamaruddin, and Thomas Mathew. Tricuspid Annular Plane Systolic Excursion: Comparing Transthoracic to Transesophageal Echocardiography. *Journal of Cardiothoracic and Vascular Anesthesia*, 31(2):590–594, April 2017.

[78] Y Le Manach and G.S. Collins. Disagreement between cardiac output measurement devices: which device is the gold standard? *British Journal of Anaesthesia*, 116(4):451–453, April 2016.

[79] Aarne Feldheiser, Peter Juhl-Olsen, Michael Nordine, Matthias Stetzuhn, Ludwig Wiegank, Fabian Knebel, Sascha Treskatsch, and Christian Berger. A comprehensive echocardiographic analysis during simulated hypovolaemia: An observational study. *Eur J Anaesthesiol*, 40(8):578–586, June 2023. Place: England.

[80] Alexander Schmeißer, Thomas Rauwolf, Thomas Groscheck, Katharina Fischbach, Siegfried Kropf, Blerim Luani, Ivan Tanev, Michael Hansen, Saskia Meißler, Kerstin Schäfer, Paul Steendijk, and Ruediger C. Braun-Dullaeus. Predictors and prognosis of right ventricular function in pulmonary hypertension due to heart failure with reduced ejection fraction. *ESC Heart Failure*, 8(4):2968–2981, 2021. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ehf2.13386.

[81] Alexander Schmeisser, Thomas Rauwolf, Thomas Groscheck, Siegfried Kropf, Blerim Luani, Ivan Tanev, Michael Hansen, Saskia Meißler, Paul Steendijk, and Ruediger C Braun-Dullaeus. Pressure–volume loop validation of TAPSE/PASP for right ventricular arterial coupling in heart failure with pulmonary hypertension. *European Heart Journal - Cardiovascular Imaging*, 22(2):168–176, November 2020. _eprint: https://academic.oup.com/ehjcimaging/article-pdf/22/2/168/35973729/jeaa285.pdf.

[82] Hongmin Zhang, Hui Lian, Xiaoting Wang, Qing Zhang, and Dawei Liu. Tricuspid annular plane systolic excursion/mitral annular plane systolic excursion ratio in critically ill patients: an index of right- and left-ventricular function mismatch and a risk factor for cardiogenic pulmonary edema. *BMC Anesthesiol*, 23(1):175, May 2023.

[83] Robert Naeije and Roberto Badagliacca. The overloaded right heart and ventricular interdependence. *Cardiovascular Research*, 113(12):1474–1485, August 2017. _eprint: https://academic.oup.com/cardiovascres/article-pdf/113/12/1474/20139140/cvx160.pdf.

[84] C. Nafati, M. Gardette, M. Leone, L. Reydellet, V. Blasco, A. Lannelongue, F. Sayagh, S. Wiramus, F. Antonini, J. Albanèse, and L. Zieleskiewicz. Use of speckle-tracking strain in preload-dependent patients, need for cautious interpretation! *Ann. Intensive Care*, 8(1):29, December 2018.

[85] Paul Glasziou, Les Irwig, and Jonathan J Deeks. When should a new test become the current reference standard? *Ann Intern Med*, 149(11):816–822, December 2008. Place: United States.

[86] Maurizio Cecconi, Andrew Rhodes, Jan Poloniecki, Giorgio Della Rocca, and R Michael Grounds. Bench-to-bedside review: The importance of the precision of the reference technique in method comparison studies – with specific reference to the measurement of cardiac output. *Crit Care*, 13(1):201, 2009.

Glossary

fps frames per second 1, 26, 40

AI artificial intelligence 6, 10, 30
ANN artificial neural network 7
ARDS acute respiratory distress syndrome 2
AUC area under the curve 5

CMR cardiac magnetic resonance 5, 14
CNN convolutional neural network 6–9, 13, 15, 16

DL deep learning 1, 6, 10, 11, 40

ED end diastole 5, 6, 11, 13–17, 21, 28
ED-RVL end-diastolic RV lenght 4, 5, 13, 37
ED-RVL$_{fw}$ free wall ED-RVL 13, 14, 18, 21, 24, 26
ED-RVL$_{mid}$ midpoint ED-RVL 13, 14, 18, 24, 26
ED-RVL$_{sep}$ septal ED-RVL 13, 14, 18, 24, 26
ES end systole 5, 6, 11, 13–17, 21
ES-RVL end-systolic RV lenght 13, 37
ES-RVL$_{fw}$ free wall end-systolic RV lenght (ES-RVL) 13, 14, 18, 21, 24, 26
ES-RVL$_{mid}$ midpoint ES-RVL 13, 14, 18, 24, 26
ES-RVL$_{sep}$ septal ES-RVL 13, 14, 18, 24, 26

FW free wall 2, 4, 5, 29

GE General Electric 13, 37
GE Vingmed Ultrasound General Electric (GE) Vingmed Ultrasound 13

ICU intensive care unit 1, 2, 5, 10, 11, 19, 30

LoA limits of agreement 10, 19, 21, 27, 28, 30
LV left ventricle 2, 5, 10, 14, 29

MAPSE mitral annular plane systolic excursion 10, 14
ME4C mid esophageal 4 chambers view 21
ML machine learning 6

NLP natural language processing 9

PAC pulmonary arterial catheter 2
PAH pulmonary arterial hypertension 4
PCA principal component analysis 6
PH pulmonary hypertension 2

RA right atrium 2, 11
ReLU rectified linear unit 7, 8
RV right ventricle 1, 2, 4–6, 10, 11, 13–17, 21, 23, 24, 26–30, 37, 39, 40
RVEDA RV end-diastolic area 5, 13, 14, 17, 18, 21, 23, 26, 28
RVEF right ventricular ejection fraction 4, 5, 11, 14
RVESA RV end-systolic area 5, 13, 14, 17, 18, 21, 23, 26, 28
RVFAC RV fractional area change 5, 6, 11, 13–15, 17, 18, 21, 23, 26, 28–30, 40
RVL RV lenght 4, 6, 13, 21, 28, 29
RVLS RV linear strain 5, 13, 21, 25, 29, 30
RVLS$_{fw}$ RV linear free wall strain 5, 14, 18, 21, 25, 26
RVLS$_{global}$ RV linear global strain 5, 13, 14, 18, 25, 26
RVLS$_{mid}$ RV linear midpoint strain 5, 14, 18, 25, 26
RVLS$_{sep}$ RV linear septal strain 5, 14, 18, 25, 26
RVLSF RV longitudinal shortening fraction 5, 6
RVS RV strain 4–6, 11, 14, 28

SOTA state of the art 8, 9
std standard deviation 1, 11, 19, 21
STE speckle tracking echocardiography 5, 6, 29

TA tricuspid annulus 1, 4, 5, 10, 11, 13–18, 21, 27, 29, 37, 40
TAD TA diameter 5, 21, 22, 27
TADD TA end-diastolic diameter 13, 14, 18, 21, 22, 26
TAPSE tricuspid annular plane systolic excursion 4–6, 11, 13, 14, 21, 29
TAPSE$_{fw}$ free wall tricuspid annular plane systolic excursion 1, 4, 13–15, 18, 19, 21, 22, 26, 27, 29, 40
TAPSE$_{global}$ global tricuspid annular plane systolic excursion 4, 13, 14, 18, 21, 22, 26, 27
TAPSE$_{sep}$ septal tricuspid annular plane systolic excursion 4, 13–15, 18, 21, 22, 26, 27
TASD TA end-systolic diameter 13, 14, 18, 21, 22, 26
TEE transesophageal echocardiography 1, 3–5, 10, 11, 19, 27, 29, 30, 38, 40
TTE transthoracic echocardiography 1, 3–5, 10, 11, 29, 30, 40
TV tricuspid valve 2, 11, 39

ViT vision transformer 6, 9

A. Application of 2D Pipeline to 3D Acquisitions

This appendix details the preliminary investigation into applying the developed 2D pipeline to 3D TEE acquisitions. The objective was to assess the generalization of the model on 4-chamber views extracted from volumetric data.

A.1. Data Acquisition

Mid-esophageal single-beat full-volume 3D TEE images focusing on the Right Ventricle (RV) were obtained from mechanically ventilated patients under general anesthesia. All data collection utilized a GE Healthcare E95 scanner equipped with a 6VT-D probe (Horten, Norway). Scanner settings were optimized to achieve a frame rate of at least 15 volumes per second for 3D acquisitions.

A.2. 3D Dataset Preprocessing

To evaluate the feasibility of applying the 2D trained model to 3D data, it was necessary to extract representative 4-chamber views from the volumetric acquisitions. The manual 3D RV segmentations provided with the dataset were utilized to define the optimal slicing plane for reconstruction. Specifically, the mesh corresponding to the first frame of each acquisition served as the reference for guiding the slicing procedure.
The plane alignment process was performed using 3D Slicer. Three specific landmark points on the tricuspid valve were identified within the 3D mesh to establish the valve's orientation(fig. 20).

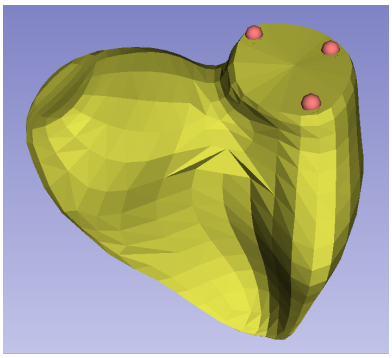


Figure 20: Visualization of the preprocessing in 3D Slicer. Coordinates of three points on the tricuspid valve were extracted from the mesh to calculate the equation of the plane parallel to the valve annulus.

These coordinates were subsequently mapped onto the acquisition's grid space to define a transformation aligning the plane formed by the three points with the grid's $x - y$ plane. Based on this alignment, an axis parallel to the z-axis and passing through the geometric center of the tricuspid valve was derived. This axis served as the center of rotation for generating a series of intersecting planes at 10° angular increments. By reslicing the volume along these planes, 18 distinct 2D slices were generated for each volumetric frame.
From this pool of generated slices, the view that most accurately reproduced the standard 4-chamber projection was manually selected, resulting in a dataset of 2D images derived directly from the 3D acquisitions. A qualitative degradation in image resolution and tissue definition is observable in the extracted slices compared to native 2D acquisitions, with some samples exhibiting significant artifacts (fig. 21).

A.3. Prediction on the Reslices Extracted from the 3D Dataset

The generalization performance of the best-performing model on the 3D-derived dataset was assessed through visual inspection. The inference results indicated a significant shift failure to the 3D domain; in the majority of cases, the predicted landmarks were incoherent, with keypoints distributed stochastically across the image rather than aligning with anatomical features.
Successful detection was observed in only a single acquisition within the test set, which was noted to possess visual characteristics most similar to the native 2D domain. Given the consistent failure of the model to adapt to the lower resolution and distinct textural properties of the 3D-derived slices, a quantitative statistical analysis was deemed unnecessary. An illustrative example of these prediction failures is presented in fig. 22.
Given these findings, further work is needed to adapt our pipeline to 3D images. This could involve, for example, fine-tuning the model on images extracted from the 3D dataset to adapt it to the different domain. Furthermore,

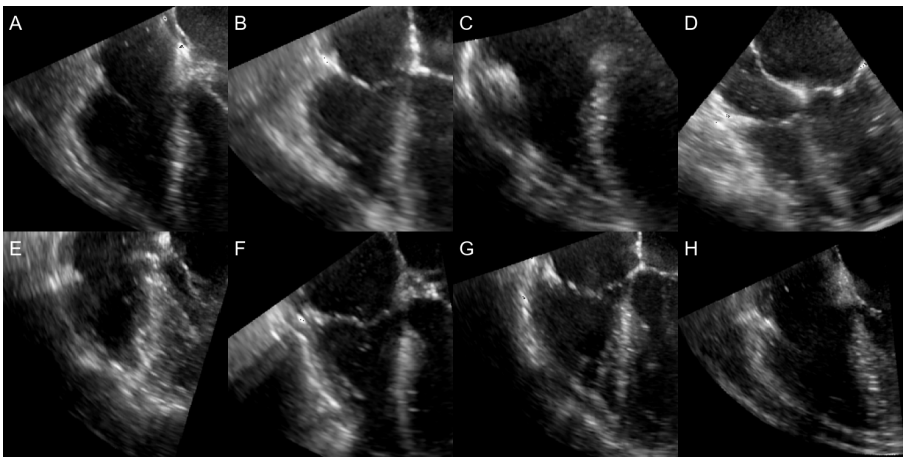


Figure 21: Examples of 2D slices extracted from the 3D volumetric acquisitions. A general reduction in image quality and resolution is evident compared to native 2D acquisitions. Several samples, such as those in panels C, E, and H, display very low quality.

a model trained to automatically extract 2D slices from 3D acquisitions, such as in [41], could be beneficial to streamline the process. Finally, a fully 3D pipeline could be implemented, for instance, to segment the TV or the RV directly.

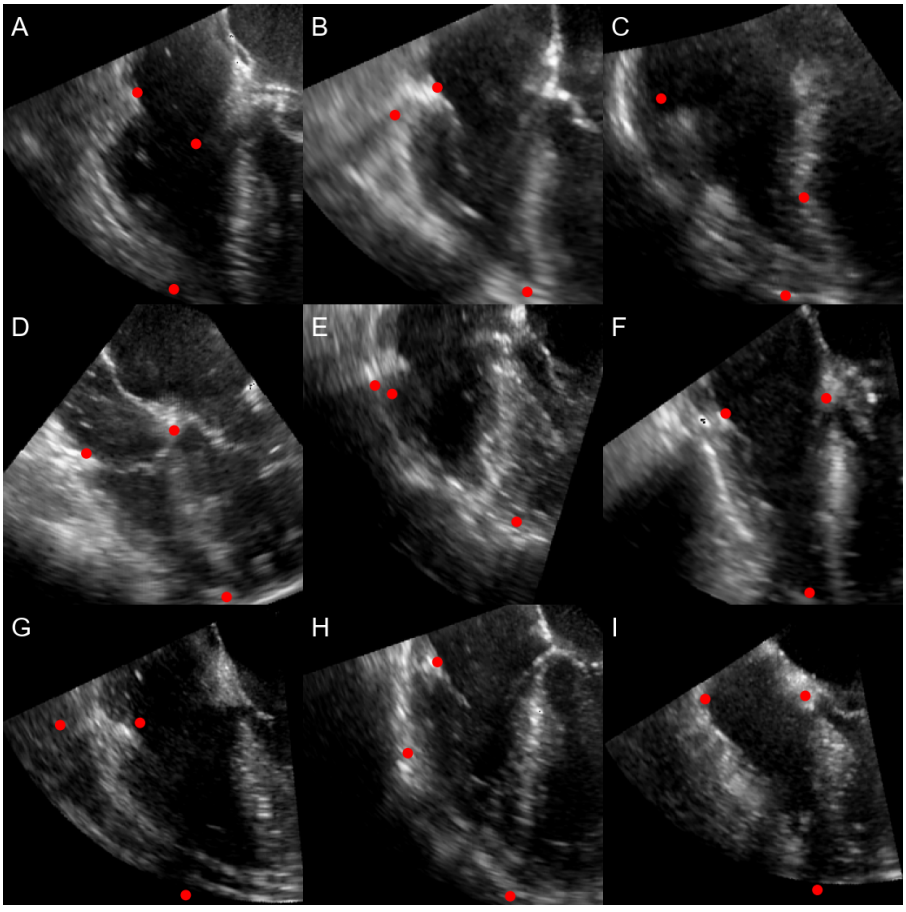


Figure 22: Representative examples of model predictions on the 3D-derived test set. The model fails to localize anatomical landmarks accurately, resulting in scattered and nonsensical keypoint placement.

Abstract in lingua italiana

La disfunzione e l'insufficienza del ventricolo destro (RV) sono complicanze rilevanti nelle terapie intensive, soprattutto nei pazienti sottoposti a ventilazione meccanica. L'ecocardiografia transesofagea (TEE) è la tecnica di imaging di riferimento, ma le attuali valutazioni si basano su analisi manuali offline, limitando il monitoraggio continuo della funzione del RV. I metodi automatizzati basati su deep learning (DL) sono stati sviluppati quasi esclusivamente per l'ecocardiografia transtoracica (TTE).
Questo lavoro presenta AutoRV, sistema completamente automatizzato basato su DL, progettato per il monitoraggio in tempo reale del RV tramite immagini TEE 2D in sezione medio-esofagea. Il modello è stato addestrato e validato su 8,424 frame provenienti da 52 pazienti ventilati. Sono state confrontate diverse architetture di reti neurali convoluzionali, tra cui varianti di ResNet, ResNeXt50, SwinUNETR e U-Net, per individuare tre landmark clinicamente rilevanti: i punti di inserzione settale e della parete libera dell'annulus tricuspidale (TA) e l'apice del RV.
Tra le architetture valutate, la U-Net ha ottenuto le migliori prestazioni, con errore medio di localizzazione pari a 4.16 ± 3.36 mm sul set di test. Il punto di inserzione del TA sulla parete libera è stato identificato con elevata precisione (2.93 ± 1.80 mm). La pipeline è computazionalmente efficiente, con velocità di inferenza di 248 frame/s (*fps*), adatta al monitoraggio continuo in tempo reale.
AutoRV calcola automaticamente gli indici clinici principali, con bias contenuto rispetto alle misurazioni manuali: 1.09 ± 2.25 mm per TAPSE_{fw} e $0.19 \pm 6.64\%$ per RVFAC. Altri parametri mostrano maggiore variabilità, soprattutto quelli dipendenti dall'apice del RV e dal punto di inserzione settale, coerentemente con le prestazioni inferiori su questi landmark.
AutoRV è il primo sistema end-to-end in grado di fornire valutazione automatizzata del RV da immagini TEE. La sua elevata velocità di inferenza e la capacità di estrarre oggettivamente parametri clinici standard ne evidenziano il potenziale per il monitoraggio cardiaco continuo in terapia intensiva.

Parole chiave: Ventricolo Destro, Ecocardiografia Transesofagea, TAPSE, RVFAC, deformazione del ventricolo destro, apprendimento profondo, Terapia Intensiva, Ultrasuoni