M(otion)-mode Based Prediction of Cardiac Function on Echocardiograms

Thomas M. Sutter[*]
ETH Zurich
thomas.sutter@inf.ethz.ch

Sebastian Balzer[*]
ETH Zurich
sebastian.balzer@inf.ethz.ch

Ece Ozkan
ETH Zurich
ece.oezkanelsen@inf.ethz.ch

Julia E. Vogt
ETH Zurich
julia.vogt@inf.ethz.ch

Abstract

Early detection of cardiac dysfunction through routine screening is vital for diagnosing cardiovascular diseases. An important metric of cardiac function is the left ventricular ejection fraction (EF), which is used to diagnose cardiomyopathy. Echocardiography is a popular diagnostic tool in cardiology, with ultrasound being a low-cost, real-time, and non-ionizing technology. However, human assessment of echocardiograms for calculating EF is both time-consuming and expertise-demanding, raising the need for an automated approach. Earlier automated works have been limited to still images or use echocardiogram videos with spatio-temporal convolutions in a complex pipeline. In this work, we propose to generate images from readily available echocardiogram videos, each image mimicking a M(otion)-mode image from a different scan line through time. We then combine different M-mode images using off-the-shelf model architectures to estimate the EF and, thus, diagnose cardiomyopathy. Our experiments show that our proposed method converges with only ten modes and is comparable to the baseline method while bypassing its cumbersome training process.

Keywords: Echocardiography · M-mode Ultrasound · Ejection Fraction

1 Introduction

Cardiovascular diseases (CVD) are the leading cause of death worldwide, responsible for nearly one-third of global deaths [1]. Accurate assessment of cardiac function and early detection of cardiac dysfunction through routine screening is essential, as clinical management and behavioural changes can prevent hospitalizations and premature deaths. A critical assessment of the cardiac function is the measurement of left ventricular (LV) ejection fraction (EF), which is expressed as a percentage and is the ratio of change in LV end-diastolic volume (EDV) and LV end-systolic volume (ESV) determined by EF = (EDV - ESV) / EDV [2].

Echocardiography is the most common and readily available diagnostic tool to assess cardiac function, ultrasound (US) imaging being a low-cost, non-ionizing, and rapid acquisition technology. However, manual assessment of echocardiograms is time-consuming, operator-dependent and expertise-demanding. Thus, there is a clear need for an automated method to assist clinicians in estimating EF. Previous automated works exploit either still-images [3–5] or spatio-temporal convolutions on B(rightness)-mode echocardiography videos [6] to predict EF. However, still-image-based methods have a high variability [7] and video-based methods rely on a complex pipeline with larger models.

[*] Shared first authorship.

36th Conference on Neural Information Processing Systems (NeurIPS 2022).

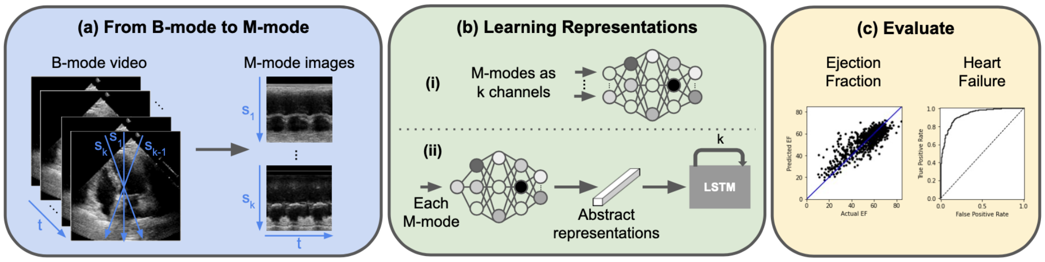


Figure 1: Overview of our proposed method. (a) Generate M-mode images from B-mode echocardiography videos at k different scan lines. (b) Combine these either (i) with early fusion by feeding them into k channels of a model architecture or (ii) with late fusion by using a model architecture to learn an abstract representation for each M-mode image and fuse them with an LSTM block to preserve their scan order. (c) Evaluate the EF prediction to diagnose cardiomyopathy.

M(otion)-mode is a form of ultrasonography in which a single scan line is emitted and received at a high frame rate through time to evaluate motion to assess different diseases and functions [8]. M-mode is often utilized in clinical practice e. g. in lung ultrasonography [9, 10] or echocardiography [11–14]. Since cardiac function assessment relies on heart dynamics, using M-mode images can be an excellent alternative to B-mode image- or video-based methods. However, little effort is directed toward exploiting M-mode images in an automated manner. A few existing works [15, 16] reconstruct M-mode images from B-mode videos to detect pneumothorax using CNNs. Furthermore, authors in [17] propose an automatic landmark localization method in M-mode images. The only more related method using M-mode images in an automated manner to estimate EF is [18], which uses single M-mode images in parasternal long-axis view to measure chamber dimensions for calculating EF.

Our contribution. In this work, we propose to extract images from readily available B-mode echocardiogram videos [7], each image mimicking an M-mode image from a different scan line of the heart. We then combine the different artificial M-mode images using off-the-shelf model architectures and estimate their EF to diagnose cardiomyopathy. This allows the model to naturally observe the motion and sample the heart from different angles while bypassing cumbersome 3D models. To the best of our knowledge, this is the first work on image-based and temporal information incorporating cardiac function prediction method to estimate EF. Furthermore, our method can easily be applied to other problems where cardiac dynamics play an essential role in the diagnosis.

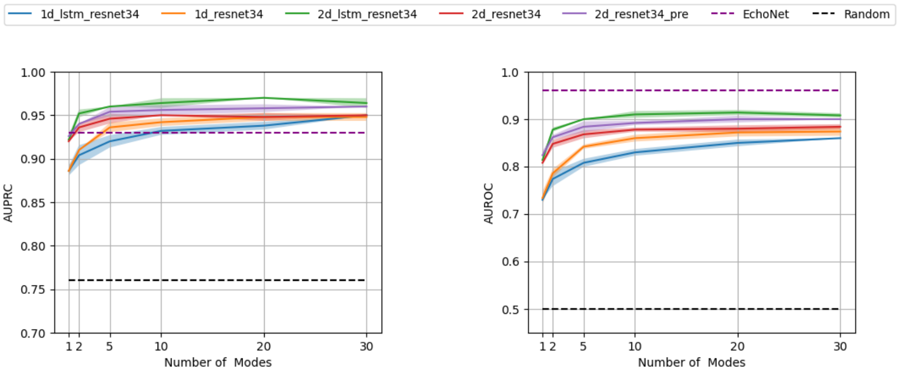
2 Method

This work aims to create a simple pipeline with as little manual intervention as possible; thus, our method consists of two parts. The first part is the extraction of M-mode images from B-mode videos, and the second is neural network learning to predict EF from M-mode images, as shown in Figure 1.

From B-mode videos to M-mode images. First, we want to generate K M-mode images from a single B-mode video, as in Figure 1(a). Assume B-mode videos are given of size $h \times w \times t$ with h being height, w width and t number of frames of the video. The k-th M-mode image is a single line of pixels through the center of the image with an angle θ_k over frames, assuming LV is around the center throughout the video. The M-mode image corresponding to θ_k is then of size $s_k \times t$, with s_k as the length of the scan line. For simplicity, we set $s_k = h \; \forall \; k$ independent of its angle θ_k. For generating multiple M-mode images, a set of K angles $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_K]$ is used to generate K M-mode images, where the angles $\boldsymbol{\theta}$ are equally spaced between 0° and 180°.

Learning representations for M-mode images. We evaluate two fusion methods for aggregating information among the K M-mode images: early-fusion and late-fusion [19], as in Figure 1(b). With early-fusion, we construct a $K \times s \times t$ image with the K M-mode images being the K channels of the newly created image. In late-fusion, we first infer an abstract representation $\boldsymbol{z}_k$ for each k-th M-mode image. The representations $\boldsymbol{z}_k$ are then aggregated using an LSTM cell [20].

Independent of the fusion principle, we utilize a standard ResNet architecture [21] with either 1D- or 2D-convolutional layers. We use 1D-convolutions when we model the t axis of the M-mode images as the temporal dimension and the s axis only as the spatial dimension. When using 2D-convolutions,



(a) AUPRC

(b) AUROC

Figure 2: Performance for different number of M-mode images K using 1D, 2D and LSTM based models. (a) evaluates the classification performance with respect to AUPRC and (b) AUROC.

we assume a single M-mode image as a 2D gray-scale image with two spatial dimensions, s and t. Additionally, we make use of models being pre-trained on ImageNet [22] when we use 2D-convolutions. We use a ResNet-34 as the model backbone independent of the remaining settings.

Training and Evaluation of Models. During training, all models optimize the estimation of EF as a regression task. For testing, we use a constant threshold τ for classifying cardiomyopathy. In all experiments, we set $\tau = 0.5$. Hence, an estimation of $\hat{\tau} < 0.5$ results in classifying a sample cardiomyopathic. All models are trained for 90 epochs using Adam optimizer [23] with an initial learning rate of 0.0001 and a batch size of 64.

3 Experiments and Results

Dataset. We use the publicly available EchoNet-Dynamic dataset [7]. The dataset contains 10030 apical-4-chamber echocardiography videos from individuals who underwent imaging between 2016 and 2018 as part of routine clinical care at Stanford University Hospital. Each B-mode video was cropped and masked to remove information outside the scanning sector, then downsampled into standardized 112×112 pixel videos. For simplicity, we used videos with at least 112 frames.

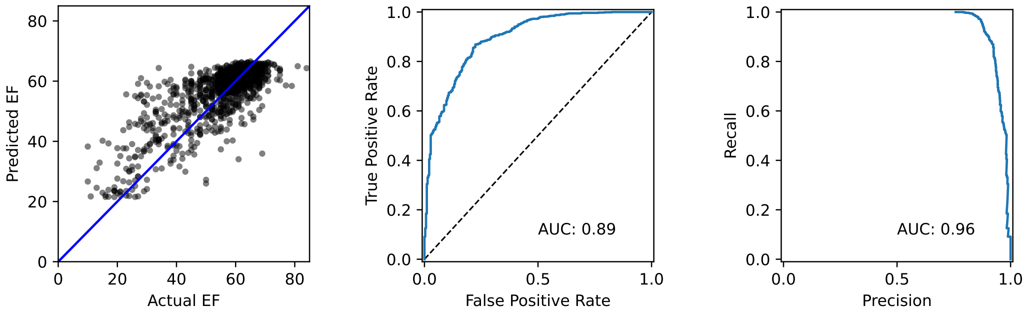
Experiments. We evaluate the performance of the model using classification accuracy for five random seeds. Figure 2 shows the performance of the different models compared to the EchoNet model. We evaluate all models using the area under the receiver operating characteristic (AUROC) and the area under the precision-recall curve (AUPRC). The black dotted line shows the performance of a random classifier classifying samples as cardiomyopathic with a probability of 0.5. The purple dotted line shows the performance of the baseline EchoNet model [6].

Our experiments show that all models benefit from an increasing number of modes K. Although the proposed work using M-mode-based information is not able to outperform the more complicated EchoNet model with respect to AUROC, all models outperform EchoNet with respect to AUPRC. The models using 2D-convolutions only need two modes on average to achieve a better performance, but with ten modes, all models outperform the baseline model using a much simpler pipeline.

Interestingly, the 2D-LSTM model even outperforms the Imagenet-pretrained model, highlighting the potential of M-mode images for automated US classification. For this, we also show predicted EF compared with the reported EF in Figure 3a, receiver-operating characteristic curve in Figure 3b and precision-recall curve in Figure 3c for the diagnosis of heart failure with reduced EF.

4 Discussion and Conclusion

In this work, we propose to generate M-mode images from readily available B-mode echocardiography videos and fuse these to estimate EF and, thus, cardiac dysfunction. Our results show that our proposed



(a) (b) (c)

Figure 3: Performance for ten M-mode images using 2D-LSTM model. (a) shows predicted EF compared to reported EF, (b) receiver-operating characteristic curve and (c) precision/recall curve for the diagnosis of cardiomyopathy.

method is comparable to the baseline method while avoiding its complex training routine and reducing the need for expensive expert input.

It is known that the conventional M-mode images have a very high sampling rate, which results in a high temporal resolution so that even very rapid motion can be recorded. Nevertheless, we do not use the M-mode images collected directly from the US machines such that there is no need for an additional data collection step. The generated M-mode images have significantly less temporal resolution than that of the conventional M-mode images from US machines. However, our results indicate that exploiting generated M-mode images does not limit the performance for EF estimation.

The dataset in this work includes human expert tracings of the LV for each echocardiogram. As a future work we plan to evaluate the effect of an informed starting axis which always cuts the left ventricle and explore self-supervised learning methods, which do not require expert labeling.

5 Potential Societal Impact

Developing methods in the medical domain offers huge potential, especially for regions where the density of skilled doctors is not high enough to provide everybody with the required medical support. Therefore, low-resource, efficient, and easy-to-use methods are even more critical. On the other hand, a model in the medical domain needs to be handled carefully, and further research is needed to ensure fairness and reliability concerning underlying and hidden attributes of patients.

References

[1] "Cardiovascular diseases (CVDs)." https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds).

[2] D. Bamira and M. H. Picard, "Imaging: EchocardiologyAssessment of Cardiac Structure and Function," in *Encyclopedia of Cardiovascular Research and Medicine*, pp. 35–54, Elsevier, 2018.

[3] A. Madani, J. R. Ong, A. Tibrewal, and M. R. K. Mofrad, "Deep echocardiography: data-efficient supervised and semi-supervised deep learning towards automated diagnosis of cardiac disease," *npj Digital Medicine*, vol. 1, no. 1, 2018.

[4] J. Zhang, S. Gajjala, P. Agrawal, G. H. Tison, L. A. Hallock, L. Beussink-Nelson, M. H. Lassen, E. Fan, M. A. Aras, C. Jordan, K. E. Fleischmann, M. Melisko, A. Qasim, S. J. Shah, R. Bajcsy, and R. C. Deo, "Fully Automated Echocardiogram Interpretation in Clinical Practice," *Circulation*, vol. 138, no. 16, pp. 1623–1635, 2018.

[5] A. Ghorbani, D. Ouyang, A. Abid, B. He, J. H. Chen, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Deep learning interpretation of echocardiograms," *npj Digital Medicine*, vol. 3, no. 1, 2020.

[6] D. Ouyang, B. He, A. Ghorbani, N. Yuan, J. Ebinger, C. P. Langlotz, P. A. Heidenreich, R. A. Harrington, D. H. Liang, E. A. Ashley, and J. Y. Zou, "Video-based AI for beat-to-beat assessment of cardiac function," *Nature*, vol. 580, no. 7802, pp. 252–256, 2020.

[7] D. Ouyang, B. He, A. Ghorbani, M. P. Lungren, E. A. Ashley, D. H. Liang, and J. Y. Zou, "Echonet-dynamic: a large new cardiac motion video data resource for medical machine learning," in *NeurIPS ML4H Workshop: Vancouver, BC, Canada*, 2019.

[8] T. Saul, S. D. Siadecki, R. Berkowitz, G. Rose, D. Matilsky, and A. Sauler, "M-Mode Ultrasound Applications for the Emergency Medicine Physician," *The Journal of Emergency Medicine*, vol. 49, no. 5, pp. 686–692, 2015.

[9] J. Avila, B. Smith, T. Mead, D. Jurma, M. Dawson, M. Mallin, and A. Dugan, "Does the Addition of M-Mode to B-Mode Ultrasound Increase the Accuracy of Identification of Lung Sliding in Traumatic Pneumothoraces?," *Journal of Ultrasound in Medicine*, vol. 37, no. 11, pp. 2681–2687, 2018.

[10] A. K. Singh, P. H. Mayo, S. Koenig, A. Talwar, and M. Narasimhan, "The Use of M-Mode Ultrasonography to Differentiate the Causes of B Lines," *Chest*, vol. 153, no. 3, pp. 689–696, 2018.

[11] R. B. Devereux, E. M. Lutas, P. N. Casale, P. Kligfield, R. R. Eisenberg, I. W. Hammond, D. H. Miller, G. Reis, M. H. Alderman, and J. H. Laragh, "Standardization of M-mode echocardiographic left ventricular anatomic measurements," *Journal of the American College of Cardiology*, vol. 4, no. 6, pp. 1222–1230, 1984.

[12] H. A. Gaspar, S. S. Morhy, A. C. Lianza, W. B. de Carvalho, J. L. Andrade, R. R. do Prado, C. Schvartsman, and A. F. Delgado, "Focused cardiac ultrasound: a training course for pediatric intensivists and emergency physicians," *BMC Medical Education*, vol. 14, no. 1, 2014.

[13] H. Skinner, H. Kamaruddin, and T. Mathew, "Tricuspid Annular Plane Systolic Excursion: Comparing Transthoracic to Transesophageal Echocardiography," *Journal of Cardiothoracic and Vascular Anesthesia*, vol. 31, no. 2, pp. 590–594, 2017.

[14] K. O. Hensel, M. Roskopf, L. Wilke, and A. Heusch, "Intraobserver and interobserver reproducibility of M-mode and B-mode acquired mitral annular plane systolic excursion (MAPSE) and its dependency on echocardiographic image quality in children," *PLOS ONE*, vol. 13, no. 5, p. e0196614, 2018.

[15] S. Kulhare, X. Zheng, C. Mehanian, C. Gregory, M. Zhu, K. Gregory, H. Xie, J. M. Jones, and B. Wilson, "Ultrasound-Based Detection of Lung Abnormalities Using Single Shot Detection Convolutional Neural Networks," in *Simulation, Image Processing, and Ultrasound Systems for Assisted Diagnosis and Navigation*, pp. 65–73, 2018.

[16] C. Mehanian, S. Kulhare, R. Millin, X. Zheng, C. Gregory, M. Zhu, H. Xie, J. Jones, J. Lazar, A. Halse, T. Graham, M. Stone, K. Gregory, and B. Wilson, "Deep Learning-Based Pneumothorax Detection in Ultrasound Videos," in *Smart Ultrasound Imaging and Perinatal, Preterm and Paediatric Image Analysis*, pp. 74–82, 2019.

[17] Y. Tian, S. Xu, L. Guo, and F. Cong, "A Periodic Frame Learning Approach for Accurate Landmark Localization in M-Mode Echocardiography," in *2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[18] P. G. Sarkar and V. Chandra, "A Novel Approach for Detecting Abnormality in Ejection Fraction Using Transthoracic Echocardiography with Deep Learning," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 16, no. 13, p. 99, 2020.

[19] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, 2018.

[20] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[21] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[22] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.

[23] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.

Appendix

Implementation Details

Our method was developed using the Python programming language (version 3.8.0) with the PyTorch deep learning library. Experiments were run on a cluster containing different NVIDIA GeForce graphic cards: GTX 1080, GTX 1080 Ti, RTX 2080 Ti.

Models and Hyperparameters

To asses the performance for different models and other hyperparameters, AUROC, AUPRC and R^2-score were used. All of these experiments were conducted using five M-mode images.

ResNet Models. ResNets [21] of varying depth – with 18, 34, 50, 101 and 152 layers – were tested to determine the model of ideal complexity for EF prediction as shown in Figure 4 and Figure 5. ResNet-18 and ResNet-34 achieve very similar results on all metrics, while the deeper ResNets perform worse overall. We selected the ResNet-34 for our experiments to have a larger model to capture sufficient information.

Hyperparameters. After choosing the model depth, other hyperparameters such as batch size, learning rate and the step size for the learning rate scheduler were investigated. We fixed the batch size at 64, learning rate at 0.0001 and the step size at 15.

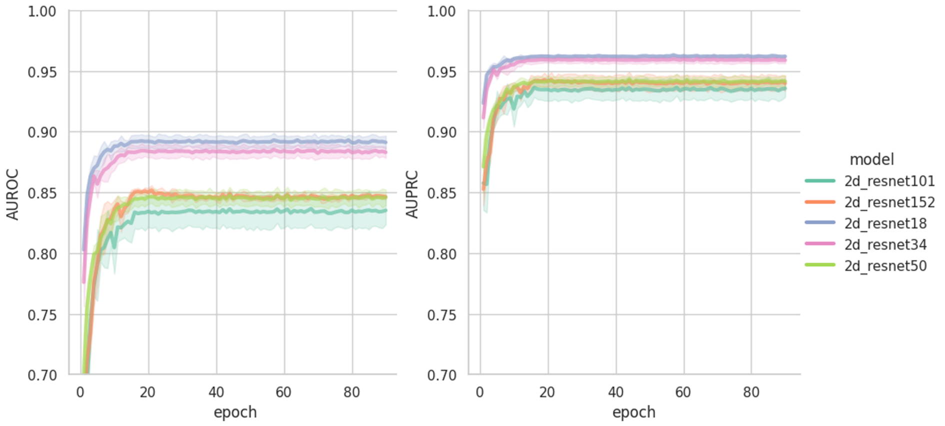


Figure 4: AUROC and AUPRC over training epochs for ResNets of varying depths.

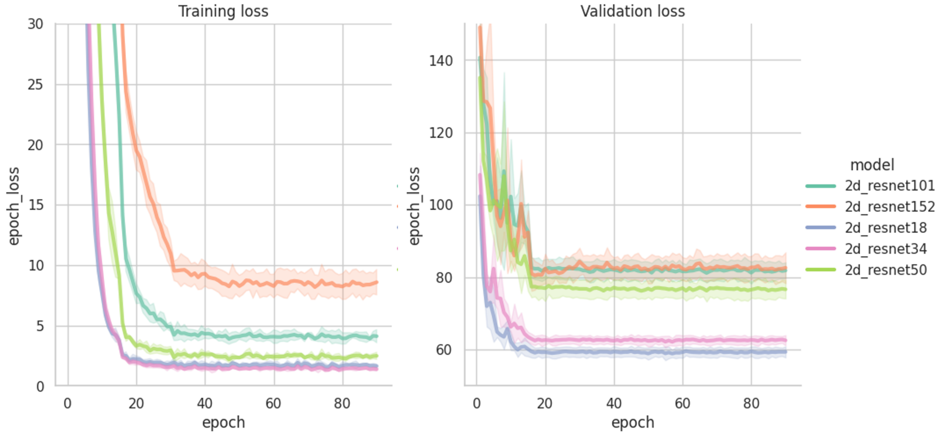


Figure 5: Training and validation loss over training epochs for ResNets of varying depths.