Deep Learning Approach for Detecting Regional Wall Motion Abnormality from Echocardiographic Images

Federico Caredda, Jose Francisco Saenz-Cogollo, Maurizio Agelli

Content Technologies & Information Management Lab, Center for Advanced Studies, Research, and Development in Sardinia, 09010 Pula - Italy.

ABSTRACT

Detecting heart motion abnormalities on echocardiographic images is crucial for an early detection of myocardial infarction in many emergency situations. However, the interpretation of the ultrasound images can be very challenging even for expert cardiologists. In this study, we propose a fully deep learning approach for detecting the presence of regional wall motion abnormalities on ultrasound sequences acquired from any of the three standard apical views used in echocardiography. We implement a two stage approach where the image sequences are first segmented by an encode-decoder network in order to isolate the left ventricular wall, and successively analyzed by a combined convolutional-recurrent network that classifies the motion of the heart muscle. The results obtained on a dataset constructed from examinations of more than 120 patients shows a very high accuracy and sensitivity.

INTRODUCTION

Regional wall motion abnormality (RWMA) refers to the occurrence of an abnormal or absent contractility of a region of the heart muscle usually involving the left ventricular (LV) wall. Since myocardial infarction or severe ischemia is the most common cause of LV wall motion abnormalities, the presence of RWMAs is one of the main indicators used in clinical cardiology for detecting a significant occult coronary artery disease not evident by electrocardiography or cardiac biomarkers (Horowitz et al. 1982). Echocardiography, with its high spatial and temporal resolution, is the best choice as a non-invasive method for assessing changes in LV wall motion. In fact, an echocardiographic assessment of RWMA is one of the recommendations of the American Heart Association and the European Society of Cardiology for the management of patients in the emergency room with symptoms of myocardial infarction without evident ischemic changes in the 12-lead ECG (Roffi et al. 2016).

Despite the recent advances in ultrasound imaging technology, detection (and quantification) of RWMA on echocardiographic images is highly subject to the observer skills, requires substantial experience, and is prone to significant inter-observer variability (Blondheim et al. 2010). Therefore, there is a need for a solution that can provide objective diagnostic assistance, particularly to the less-experienced cardiologist.

In recent years, machine learning has emerged as an effective and accurate technology for analyzing ultrasound images in order to provide support for diagnosis and/or assessment, as well as image-guided interventions/therapy (Brattain et al. 2018). Deep learning (DL), which is a branch of machine learning, is considered one of the more promising approaches because it can be used to directly process and automatically learn mid-level and high-level abstract features acquired from ultrasound images (Liu et al. 2019).

In the context of echocardiography, most of the works that have reported the use of DL techniques have been focused on the task of image segmentation mainly with the aim of delineating the left ventricular endocardium (Chen et al. 2020). A particularly relevant recent work was presented in (Leclerc et al. 2019) where a set of DL and non-DL algorithms are investigated for the task of 2D echocardiographic segmentation on a new publicly-available and fully annotated dataset called CAMUS (Cardiac Acquisitions for Multi-structure Ultrasound Segmentation). This study confirms the U-Net architecture (Ronneberger, Fischer, and Brox 2015) as one of the best approaches for biomedical image segmentation tasks.

It's only very recently that DL has been applied to the specific task of RWMA assessment. In (Degerli et al. 2020), an U-Net model is used for segmenting the LV wall in a three-phase approach that includes manual feature engineering and a final machine-learning (using a support vector machine) classification stage for detecting the presence of RWMAs. In (Kusunose et al. 2020), researchers explore the application of five different DL architectures (ResNet, DenseNet, Inception-ResNet, Inception, and Xception) to two RWMA assessment tasks (presence of RWMAs and arterial territory of RWMAs) using short-axis view images. By comparing the performance of DL models to that of experienced and less experienced echocardiographic readers, the authors found that the performance of the best model (ResNet) was comparable to that of experienced cardiologist and sonographer readers and significantly higher than that of resident readers.

In the present work, we propose a two stage approach for detecting the presence of RWMAs. First, the area of myocardium corresponding to the LV wall is segmented and extracted from the echocardiographic sequences using a DL model based on the U-Net architecture. Then, the resulting masked sequences are processed by a combined network of convolutional and recurrent modules based on a MobileNet (Howard et al. 2017) and a Gated Recurrent Unit (GRU) (Chung et al. 2014) which outputs the classification result. An overview of the proposed approach is shown in Fig 1.

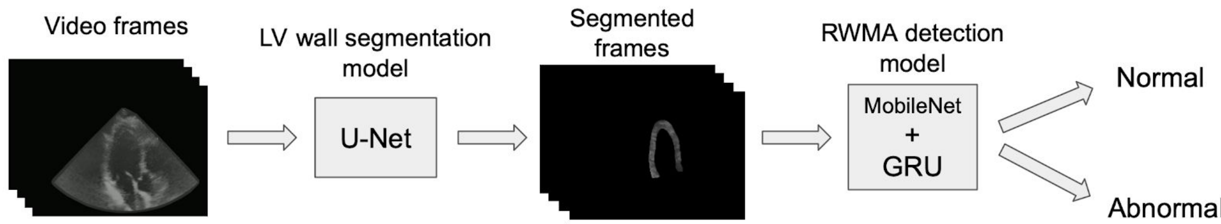

Fig. 1 Overview of the proposed framework. Echocardiographic videos are first segmented by a deep encoder-decoder network based on U-Net and successively analyzed by a combined network of convolutional and recurrent modules that classify the sequences as having normal or abnormal wall motion.

METHODS

RWMA detection dataset. The dataset used for training and validating the classification model was constructed from 288 echocardiographic examinations of 133 patients obtained in the Luigi Sacco Hospital of Milan (Italy) between May and August 2020. For each patient one or more examinations from one of the three standard apical views (A2C, A3C, and A4C) were recorded. Each examination resulted in a DICOM (Digital Imaging and Communications in Medicine) file that was parsed and analyzed to extract the area of interest in the images, saving the resulting sequence in a video file that was subsequently used for the experiments. All videos have a frame rate of 20 fps with an average duration of 3.75 s (75 frames). In this study, we use a subset of 185 video sequences from 124 patients excluding those where the entire LV wall was not visible for an entire cardiac cycle. The composition of the dataset is shown in Table I.

Table I. Composition of the dataset constructed for detecting the presence of RWMA

	Non-RWMA	RWMA	Total
Patients	61	63	124
Total video sequences	88	97	185
A4C video sequences	44	33	77
A2C video sequences	23	45	68
A3C video sequences	21	19	40

Segmentation dataset. High quality pixel-wise LV wall annotation of echocardiographic sequences is cumbersome and requires a large amount of expert time due to the number of images involved. Therefore, we leverage the recent publication of the high quality CAMUS dataset for training the segmentation model (Leclerc et al. 2019). The publicly available train dataset consists of annotated echocardiographic examinations of 450 patients totaling 1680 annotated images obtained in the University Hospital of St Etienne (France). The images in this dataset were optimized for measuring left ventricular ejection fraction (LVEF), the primary indicator for evaluating cardiac pump function. For each patient there are four annotated images: two images for the end-diastole in both apical two-chamber (A2C) and apical four-chamber (A4C) views, and 2 the end-systole in both A2C and A4C. For each image each pixel is annotated with a number to denote the type of structure it belongs to: background, LV myocardium, LV, and left atrium. In this work we use only the labels that indicate the area of LV myocardium.

LV wall segmentation model. We implement an encoder-decoder convolutional neural network based on the popular U-Net architecture for generating a segmentation mask for the LV wall. Given the wide range of possible U-Net designs, we select the network parameters in order to resemble the model labeled as U-Net 1 by (Leclerc et al. 2019) which demonstrated a very good performance segmenting echocardiographic images. Briefly i) the network was implemented with a depth of six; ii) the number of filters per convolutional layers increases and decreases linearly from an initial kernel size of 32; iii) 2*2 max-pooling was used as downsampling with bilinear upsampling; iv) ReLu activations were used; v) segmentation was obtained from a final softmax layer for two classes. Table II summarizes the U-Net implementation used here.

Table II. Architecture of the LV wall segmentation network based on the U-Net implementation proposed by (Leclerc et al. 2019).

Level	Layer	Kernel / Pool size	Activation	Connection
D1	Conv	32 (3,3)	ReLU	*
	Conv	32 (3,3)	ReLU	
	MaxPooling	(2*2)		
D2	Conv	32 (3,3)	ReLU	**
	Conv	32 (3,3)	ReLU	
	MaxPooling	(2*2)		
D3	Conv	64 (3,3)	ReLU	***
	Conv	64 (3,3)	ReLU	
	MaxPooling	(2*2)		
D4	Conv	128 (3,3)	ReLU	****
	Conv	128 (3,3)	ReLU	
	MaxPooling	(2*2)		
D5	Conv	128 (3,3)	ReLU	*****
	Conv	128 (3,3)	ReLU	
	MaxPooling	(2*2)		
D6	Conv	128 (3,3)	ReLU	
	Conv	128 (3,3)	ReLU	
U1	UpSampling	(2,2)		*****
	Conv	128 (3,3)	ReLU	
	Conv	128 (3,3)	ReLU	
U2	UpSampling	(2,2)		****
	Conv	128 (3,3)	ReLU	
	Conv	128 (3,3)	ReLU	
U3	UpSampling	(2,2)		***
	Conv	64 (3,3)	ReLU	
	Conv	64 (3,3)	ReLU	
U4	UpSampling	(2,2)		**
	Conv	32 (3,3)	ReLU	
	Conv	32 (3,3)	ReLU	
U5	UpSampling	(2,2)		*
	Conv	16 (3,3)	ReLU	
	Conv	16 (3,3)	ReLU	
Seg	Conv	4 (1,1)	Softmax	

RWMA detection model. In order to analyze the motion of the LV ventricular wall, we implement a network based on a Gated Recurrent Unit (GRU) with a MobileNet backbone and fully connected decision layers. The MobileNet backbone was pre-trained with ImageNet and provides a fast but robust convolutional core, whereas the GRU provides the temporal analysis power with efficient memory usage. The GRU is like a long short-term memory (LSTM) with a forget gate, but has fewer parameters than LSTM, as it lacks an output gate (Chung et al. 2014). This choice follows an approach similar to that used for action recognition models, as the problem treated falls within the case history of kinetic problems. The performances regarding the training time and the model accuracy obtained reach an optimal result after 100 epochs. The following choices were made to obtain the best results on our dataset: i) only the last 9 layers of the pretrained MobileNet are trainable; ii) the size of GRU layer output-space is set to 64 units; iii) in order to analyze at least one complete cardiac cycle on each input sequence, the GRU input sequence size is set to 25 frames, which (given a frame rate of 20 fps) can cover an entire cardiac cycle even for a patient with a heart rate as slow as 50 beats per minute; iv) frames are resized to 224 x 224 so that the input of the model has a size of (25, 224, 224, 3) ; v) within the GRU layer Sigmoid activations in used, while we use ReLu activations on 4 subsequent decision layers, with output space sizes of 1024, 512, 128 and 64 respectively; vi) a dropout of 50% is used so as to avoid data overfitting; vii) at the end, classification is obtained from a final softmax layer for two classes.

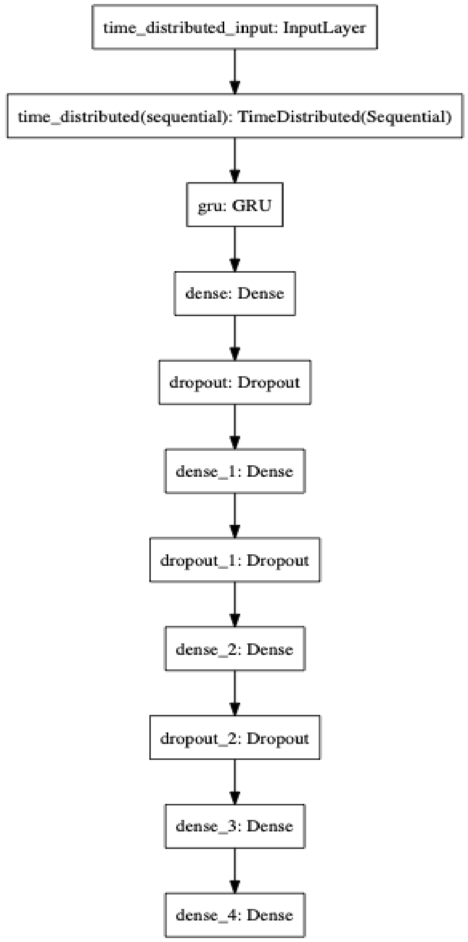

Fig. 2. Architecture of the proposed RWMA detection network.

RESULTS

LV wall segmentation experiments. The segmentation model was implemented in Python using Pytorch 1.6. and trained on a subset of the CAMUS train dataset that consists of 1620 images (405 patients), leaving the rest of the dataset (180 images from 45 patients) for evaluating the performance of the model. As preprocessing, images were resized to 256x256 pixels before performing density normalization, and padding was applied before the 3*3 convolutions. During training a batch size of 8 was used, the Adam method was used for stochastic optimization, and the number of epochs was set to 60. In order to improve the inference power of the model, a scheme of data augmentation based on random affine without increasing the amount of training data was adopted. In brief, each training sample was randomly rotated between -15 and 15 degrees, randomly translated between 0% and 10% of the image dimension, and randomly scaled between 50% and 100% of the original size.

The performance of the model in predicting the LV segmentation masks is evaluated in terms of pixel-wise F1-score, and the accuracy is estimated using the Dice similarity index. The Dice index gives a value between 0 (no overlap) and 1 (full overlap). The results obtained with the CAMUS dataset are a Dice index of 0.864 ± 0.043 and an F1-score of 0.866 ± 0.06. In order to verify that the resulting model is capable of inferring LV segmentation masks on the RWMA detection dataset (which was acquired in a different context), we visually inspect the segmentation masks predicted by the model on a small set of representative images and confirm the quality of the results as shown on Figure 3.

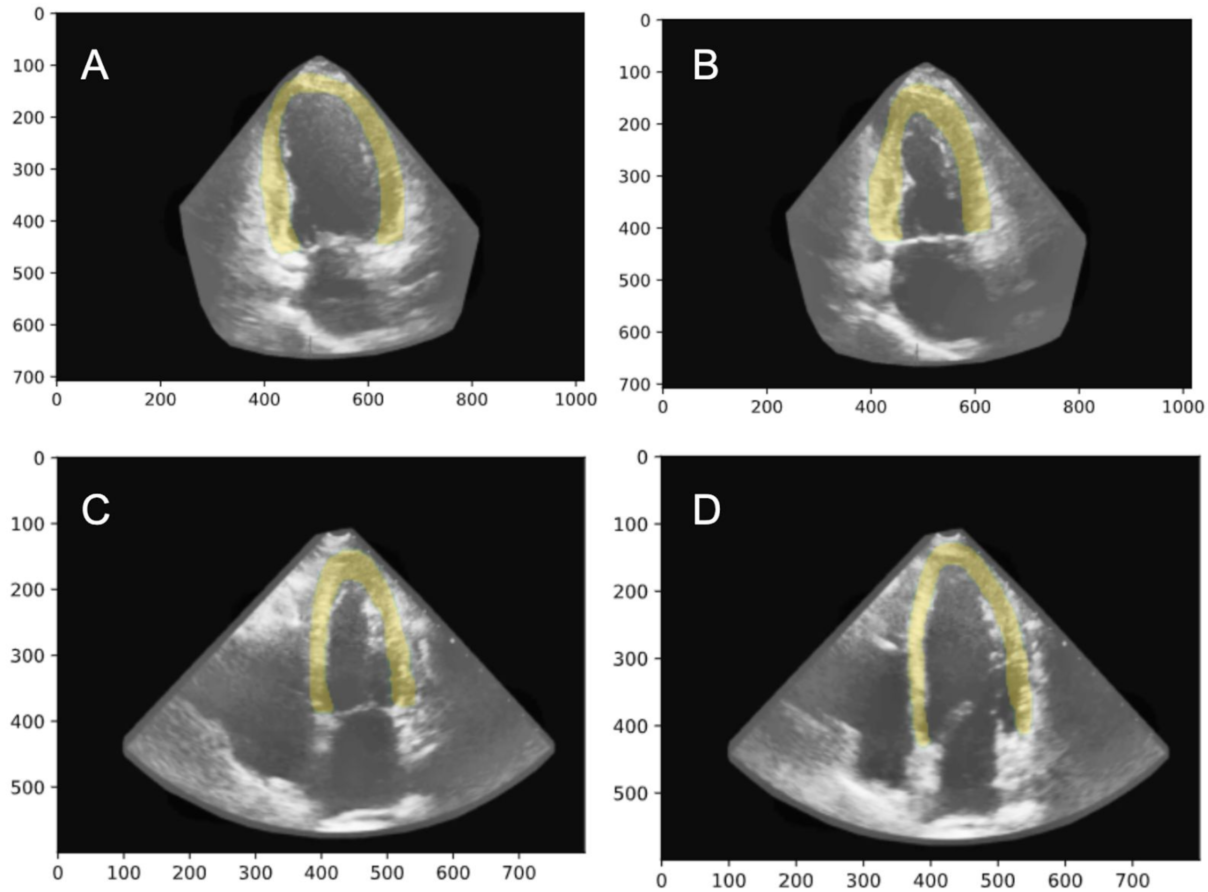

Figure 3. Masks generated by the LV segmentation model for four images in the RWMA detection dataset. A and B are A2C views obtained from a patient without RWMAs at end-diastole and end-systole respectively. C and D are A4C views obtained from a patient with RWMAs at end-diastole and end-systole respectively.

RWMA detection experiments. The RWMA detection model was also implemented in Python but this time using Keras 2.4.3 with Tensorflow 2.3.0 as backend. For the experiments, the RWMA detection dataset was used for training and validation using a k-fold cross validation (CV) scheme. Specifically, we perform a 5-fold CV where the model was trained and validated 5 times, choosing randomized and mutually exclusive validation sets each time. Videos were processed using a sliding window approach where a window of 25 frames is moved over the data, frame by frame, and the prediction is computed over the sequence inside the window. In this way, at each time step a decision is performed using the current frame and the 24 previous frames. Since for the first 24 frames the window does not have enough data yet, for a video with L frames the total number of decisions is L - 24.

During training a batch size of 8 is used, the SGD method, with a learning rate of 0.01 for each epoch, was used for stochastic optimization, and the number of epochs is set to 100. In order to improve the inference power of the model, a scheme of data augmentation based on random affine without increasing the amount of training data was adopted. In brief, each training sample, made by 25 images, was randomly rotated between -8 and 8 degrees, randomly shifted between -20% and +20% of the image dimension, and randomly zoomed in/out between 90% and 110% of the original size.

Table III shows the RWMA detection results in terms of positive predictive value (PPV), sensitivity, specificity and accuracy for each 5-fold CV together with their mean and standard deviation. The results indicate that the proposed approach achieves high detection accuracy with a mean value close to 93%. Especially remarkable is the equally high sensitivity obtained which is important in the management of patients with symptoms of myocardial infarction where false negatives may preclude a timely therapy that may save lives.

Table III. Classification metrics obtained during with a 5-fold cross validation of the proposed approach on the RWMA detection dataset.

	PPV (%)	**Sensitivity (%)**	**Specificity (%)**	**Accuracy (%)**

Fold-1	100.00	70.27	100.00	82.81
Fold-2	93.28	99.11	92.86	95.98
Fold-3	98.61	97.93	98.48	98.19
Fold-4	87.60	99.12	85.45	92.41
Fold-5	93.49	98.29	92.52	95.54
Mean	94.59	92.94	93.86	92.99
Std	4.41	11.34	5.15	5.41

Unfortunately, due to the lack of publicly available datasets and few related works it is difficult to directly compare our results with similar approaches based on deep learning. In relation to the work of (Degerli et al. 2020) we obtain metrics 10% higher with a database similar in size but with more diverse views. Our results in terms of sensitivity and specificity are more similar to those recently obtained by (Kusunose et al. 2020) though that work do not use the apical views but only the short-axis view, and only high quality images were included in the dataset.

CONCLUSIONS

In this paper we address the task of automatic detection of heart wall motion abnormalities on echocardiographic sequences. For this purpose, we proposed a novel fully deep learning approach where convolutional neural networks are used both for segmenting the area of interest in the images and for analyzing the motion of the heart muscle. An encoder-decoder network based on the U-Net model was used for isolating the left ventricular wall whereas a combined network of convolutional and recurrent modules based on MobileNet and GRU was used for classifying the motion of the segmented myocardium. Results suggest that the proposed framework can be used to detect the presence of RWMA on any of the three standard apical views used in echocardiography with high accuracy and high sensitivity.

REFERENCES

Blondheim, David S., Ronen Beeri, Micha S. Feinberg, Mordehay Vaturi, Sarah Shimoni, Wolfgang Fehske, Alik Sagie, et al. 2010. "Reliability of Visual Assessment of Global and Segmental Left Ventricular Function: A Multicenter Study by the Israeli Echocardiography Research Group." *Journal of the American Society of Echocardiography*. https://doi.org/10.1016/j.echo.2009.12.020.

Brattain, Laura J., Brian A. Telfer, Manish Dhyani, Joseph R. Grajo, and Anthony E. Samir. 2018. "Machine Learning for Medical Ultrasound: Status, Methods, and Future Opportunities." *Abdominal Radiology (New York)* 43 (4): 786–99.

Chen, Chen Chen, Chen Qin, Huaqi Qiu, Giacomo Tarroni, Jinming Duan, Wenjia Bai, and Daniel Rueckert. 2020. "Deep Learning for Cardiac Image Segmentation: A Review." *Frontiers in Cardiovascular Medicine*. https://doi.org/10.3389/fcvm.2020.00025.

Chung, Junyoung, Caglar Gulcehre, Kyunghyun Cho, and Yoshua Bengio. 2014. "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling." http://arxiv.org/abs/1412.3555.

Degerli, Aysen, Morteza Zabihi, Serkan Kiranyaz, Tahir Hamid, Rashid Mazhar, Ridha Hamila, and Moncef Gabbouj. 2020. "Early Detection of Myocardial Infarction in Low-Quality Echocardiography." http://arxiv.org/abs/2010.02281.

Horowitz, R. S., J. Morganroth, C. Parrotto, C. C. Chen, J. Soffer, and F. J. Pauletto. 1982. "Immediate Diagnosis of Acute Myocardial Infarction by Two-Dimensional Echocardiography." *Circulation*. https://doi.org/10.1161/01.cir.65.2.323.

Howard, Andrew G., Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." http://arxiv.org/abs/1704.04861.

Kusunose, Kenya, Takashi Abe, Akihiro Haga, Daiju Fukuda, Hirotsugu Yamada, Masafumi Harada, and Masataka Sata. 2020. "A Deep Learning Approach for Assessment of Regional Wall Motion Abnormality From Echocardiographic Images." *JACC. Cardiovascular Imaging* 13 (2 Pt 1): 374–81.

Leclerc, Sarah, Erik Smistad, Joao Pedrosa, Andreas Ostvik, Frederic Cervenansky, Florian Espinosa, Torvald Espeland, et al. 2019. "Deep Learning for Segmentation Using an Open Large-Scale Dataset in 2D Echocardiography." *IEEE Transactions on Medical Imaging* 38 (9): 2198–2210.

Liu, Shengfeng, Yi Wang, Xin Yang, Baiying Lei, Li Liu, Shawn Xiang Li, Dong Ni, and Tianfu Wang. 2019. "Deep Learning in Medical Ultrasound Analysis: A Review." *Engineering*. https://doi.org/10.1016/j.eng.2018.11.020.

Roffi, Marco, Carlo Patrono, Jean-Philippe Collet, Christian Mueller, Marco Valgimigli, Felicita Andreotti, Jeroen J. Bax, et al. 2016. "2015 ESC Guidelines for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation: Task Force for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-Segment Elevation of the European Society of Cardiology (ESC)." *European Heart Journal* 37 (3): 267–315.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." *Lecture Notes in Computer Science*. https://doi.org/10.1007/978-3-319-24574-4_28.