Machine Learning for Quality Assurance of Myocardial Strain Curves — Source link ↗

Mahdi Tabassian, Olivia ZulaicaIglesias, Serkan Ünlü, Jens-Uwe Voigt ...+1 more authors

Institutions: Katholieke Universiteit Leuven

Published on: 01 Oct 2018 - Internaltional Ultrasonics Symposium

Related papers:

- Assessment and Evaluation of cancer CT images using deep learning Techniques

Machine Learning for Quality Assurance of Myocardial Strain Curves

Mahdi Tabassian*, Olivia Zulaica Iglesias*, Serkan Ünlü†, Jens-Uwe Voigt*†, and Jan D'hooge*

* Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

† Department of Cardiology, University Hospitals Leuven, Leuven, Belgium

mahdi.tabassian@kuleuven.be

Abstract—Speckle tracking echocardiography (STE) is a well-established technique to quantify regional myocardial function. Reliability of STE-derived strain curves, however, depends strongly on the quality of the acquired B-mode images and can significantly be influenced by image artifacts. Artifactual images could lead to tracking errors and as a result, the measured deformation patterns might be similar to those obtained in pathology. It would thus be clinically very relevant to get feedback on the reliability (i.e. quality) of the extracted strain curves. As such, the aim of this study was to examine the utility of machine learning in the identification of artifactual strain curves. Our proposed learning framework was built upon a data imputation algorithm in order to facilitate the task of classifying (non-)artifactual curves. The obtained results confirmed the feasibility of automatic quality assurance of the STE-derived strain curves via machine learning.

I. INTRODUCTION

Two-dimensional speckle-tracking echocardiography (STE) [1] has become the modality of choice to assess regional myocardial deformation. In order to generate strain curves that well represent myocardial function, however, the STE algorithm should be provided with B-mode images of high quality. Poor image quality which can be due to artifacts, noise or clutter may result in strain curves that are not physiologically meaningful (i.e. artifactual curves). In addition, shadowing caused by ribs or lung tissue might prevent acquiring images of all myocardial segments resulting in missing strain curves.

In order to address the latter problem, we recently suggested to use a data imputation technique [2] for estimating the missing strain curves based on those that could be measured from the rest of left ventricular (LV) segments [3]. However, the former problem remains unresolved and identification of artifactual measurements is an open problem. Although this task could be done by visual reading, reliability of the assigned labels to the strain curves is strongly dependent on expertise of a cardiologist in reading this curves as well as wall motion B-mode images. Moreover, readings of expert cardiologists are subjective, time consuming and intra-/inter-observer variable.

To cope with this problem, a supervised learning approach has recently been proposed in [4] to automatically identify non-physiological strain curves by training a classifier with a group of STE-derived (non-)artifactual curves in order to learn their major temporal patterns. Although it was shown that such a learning system is capable to identify non-physiological strain curves with a good accuracy, direct analysis of the strain curves for categorizing artifactual from non-artifactual curves remains challenging given that the patterns of the artifactual strain curves might be similar to those measured from pathological LV segments.

With this in mind, we propose in this study a new machine learning system for the quality assurance of the STE-derived strain curves by taking advantage of our proposed imputation approach for aiding a classification system to better discriminate between the non-artifactual and artifactual curves.

The rest of this paper is organized as follows. Section II describes the details of data acquisition and the proposed machine learning framework. Results are presented in Section III and the obtained results are discussed in Section IV. Finally, Section V draws conclusions and summarizes the paper.

II. MATERIALS AND METHODS

A. Study Population

In the context of the ASE/EACVI task force on STE standardization [5], a group of 57 patients was selected from hospital records based on the following criteria: 1) age >18 years and ability to consent, walk, and lie in supine position for two hours; 2) a documented myocardial infarction (MI) within maximum two years before the study; 3) good acoustic window and regular heart rhythm; and 4) the presence of a late gadolinium enhancement (LGE) cardiac magnetic resonance study performed after the MI. Since not all invited patients could present for the study, five healthy volunteers were recruited as gap fillers in stand-by from the colleagues of our laboratory. The study was approved by the ethics commission of the University Hospitals Leuven, and all subjects gave written informed consent.

B. Echocardiographic Strain Imaging

Three consecutive cardiac cycles from the apical 4- and 2-chamber views of the patients were obtained during breath hold using a Vivid E9 echocardiographic machine (GE Vingmed Ultrasound, Horten, Norway). Strain measurements were performed using GE's speckle tracking software (EchoPac v20.1) based on the 18-segment model of the LV [6] (i.e. 6 segments per apical view).

An expert cardiologist assessed the tracking quality of each segment by visually comparing the tracking result with the underlying myocardial motion and categorized its corresponding strain curve as artifactual or not. After the visual assessment,
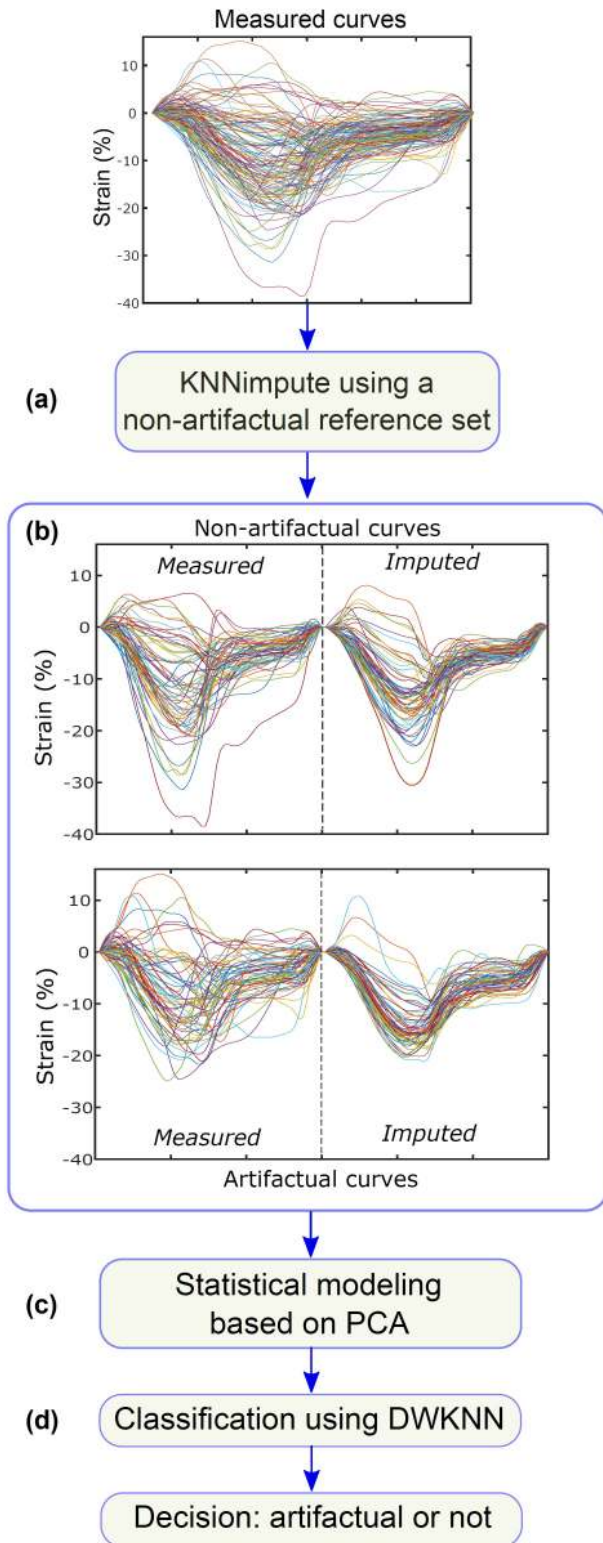
Fig. 1. Structure of the proposed machine learning framework to assess the quality of the segmental strain curves. After imputing the measured curves (a), the imputed curves and their corresponding measured curves were concatenated (b) to highlight their (dis)similarities. The concatenated curves were then statistically modeled (c) and the extracted features from them finally used to classify the strain curves into artifactual or not (d).

34 of the 57 subjects in the database were labeled as artifactual (i.e. $\approx 60\%$) in that they showed an artifactual strain curve in at least one LV segment. Around 15% of the segmental strain curves of these 34 subjects were artifactual which formed $\approx 10\%$ of all strain curves in the database.

C. Preprocessing

In order to account for the differences in the number of time points of the strain curves, which could be due to differences in frame rate and/or heart rate, a temporal alignment based on linear interpolation was performed. Systolic and diastolic phases were interpolated separately by considering the average number of points for that phase over all subjects. The segmental curves had 22 ± 3 and 39 ± 10 temporal instances during systole and diastole, respectively. After the temporal alignment, all segmental curves had 22 systolic and 39 diastolic, i.e. in total 61, time points.

D. Strain Curve Imputation

The core of the proposed machine learning framework (Fig. 1) is to use a data imputation technique [2]. The idea is that for testing the quality of a given strain curve, it can be considered missing and then be imputed using a set of non-artifactual strain curves. The measured curve and its imputed version can afterwards be contrasted in order to examine their similarity. Our hypothesis is that similarity between the patterns of an artifactual curve and its imputed version is less than that of a non-artifactual curve and its imputed trace given that the reference set used for imputation is only composed of non-artifactual curves (see Fig. 1(b)). This difference between the similarities of the measured and imputed curves of the non-artifactual and artifactual strain data can be assessed by a classifier to automatically categorize the strain curves of these two groups.

We used the k-nearest neighbor imputation (KNNimpute) technique [7] in the structure of our proposed framework (Fig. 1(a)) given its promising performance in imputing echocardiographic strain (rate) curves as shown previously in [3]. In order to impute the segmental strain curves of each subject, a group of subjects without artifactual curves was taken from the rest of the subject in the database to serve as the reference imputation set. As suggested in [3], all segmental strain curves of each subject (i.e. 12 curves of the apical 4- and 2-chamber views) were concatenated in a fixed order and imputation was performed separately for each curve of a given subject by removing the corresponding curves in the reference set.

The examined KNN values were taken from the interval of [1, 9] where the reason for testing small KNNs was the fact that by using large values, the imputed curves become too smooth [3].

E. Statistical Modeling

After performing the imputation phase, the measured and imputed segmental curves were concatenated (Fig.1(b)) and *Principal Component Analysis* (PCA) [8] was used to statistically model their major (dis)similarity patterns (Fig.1(c)),

called principal components (PCs). A subset of the most important PCs were preserved and the concatenated curves were projected onto the space spanned by these PCs to obtain a set of features required for building a classifier. The examined PCs were taken from the interval of [1, 40].

F. Automatic Classification

The distance-weighted *k*-nearest neighbor (DWKNN) method [9] was used as the classifier in our experiments (Fig.1(d)). The features extracted via PCA modeling were given to DWKNN in order to discriminate artifactual from non-artifactual strain curves by taking KNNs from the [1, 70] interval.

G. Parameter Settings with Cross-Validation

We used cross-validation (CV) to tune the parameters of the different modules of the proposed framework. To find the optimal number of KNNs for imputation, the *leave-one-patient-out* CV technique was used. The reason for adopting this CV approach was the need for using as many subjects as possible for creating the reference imputation set to have well-estimated strain curves. The *Monte Carlo* CV technique, however, was used to find the optimal number of PCs for the statistical modeling and KNNs for the automatic classification by randomly sampling 70% of the subjects in the training set and the remaining 30% in the testing set and repeating this process 200 times. This way of CV allowed us to use the subjects of both artifactual and non-artifactual classes in the testing set and examine the performance of the proposed approach many times (i.e. more than the number of subjects in the database) to reduce variance in the estimations.

III. RESULTS

Fig. 2 shows the obtained classification results using the optimal parameters for imputation (KNN = 3), statistical modeling (PC = 3) and classification (KNN = 40). Performing receiver operating characteristic (ROC) analysis in each round of the CV process, the optimal parameters were considered to be those which resulted in the best average area under the curve (AUC). Accuracy, sensitivity and specificity rates were then obtained by selecting optimal cutoff points on the ROC curves.

The proposed machine learning framework yielded an average AUC of $70 \pm 6\%$ and accuracy, sensitivity and specificity rates of $64 \pm 10\%$, $64 \pm 15\%$ and $64 \pm 12\%$, respectively.

IV. DISCUSSION

This study sought to investigate the feasibility of automatic quality assurance of the STE-derived strain curves via machine learning where the key idea was to use a data imputation technique to aid discriminating the non-artifactual curves from the artifactual ones. Considering the following practical challenges that the proposed machine learning approach had to face, its performance was promising:

- providing a reliable ground-truth for training the learning algorithm was difficult given that the cardiologist's decisions in assigning (non-)artifactual labels to the strain
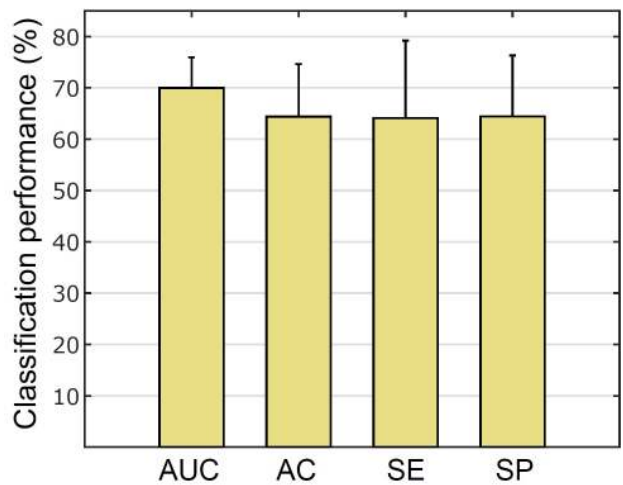

Fig. 2. Area under the curve (AUC) of the DWKNN classifier (mean and standard deviation) obtained with the optimal parameters of the proposed machine learning approach. The corresponding accuracy (AC), sensitivity (SE) and specificity (SP) rates were computed by selecting optimal cutoff points on the ROC curves.

curves were subjective and could also suffer from the intra- and inter-observer variability;
- the learning algorithm was trained using a small set of artifactual strain curves (i.e. $\approx 10\%$ of the database) since in practice, the majority of the measured strain curves are non-artifactual;
- the patterns of the artifactual curves can be similar to those of curves extracted from infarcted/diseased segments.

While the first two challenges might be addressed by carefully reading the strain curves by one or more expert cardiologists and collecting a larger set of artifactual strain curves, the latter is more difficult to be solved.

In order to better understand the difficulty of discriminating the artifactual strain curves from the non-artifactual traces, examples of the measured and imputed strain curves are shown in Fig. 3. It can be seen that the patterns of the artifactual curves measured from healthy segments are similar to those of the non-artifactual curves measured in infarcted segments. The same also holds true for the artifactual and non-artifactual curves measured from the infarcted segments. Despite these similarities, however, the idea of concatenating the measured and imputed curves appeared to be useful in making these curves distinguishable. Another observation that can be made is that the imputed curves of the non-artifactual group (either healthy or infarcted) mimicked the patterns of their corresponding measured curves while the patterns of the measured and imputed artifactual curves are more distinct. This is intuitive and expected given that the reference imputation set was only composed of the non-artifactual curves.

V. CONCLUSIONS

In this study, we have proposed a machine learning algorithm for automatic identification of the STE-derived ar-
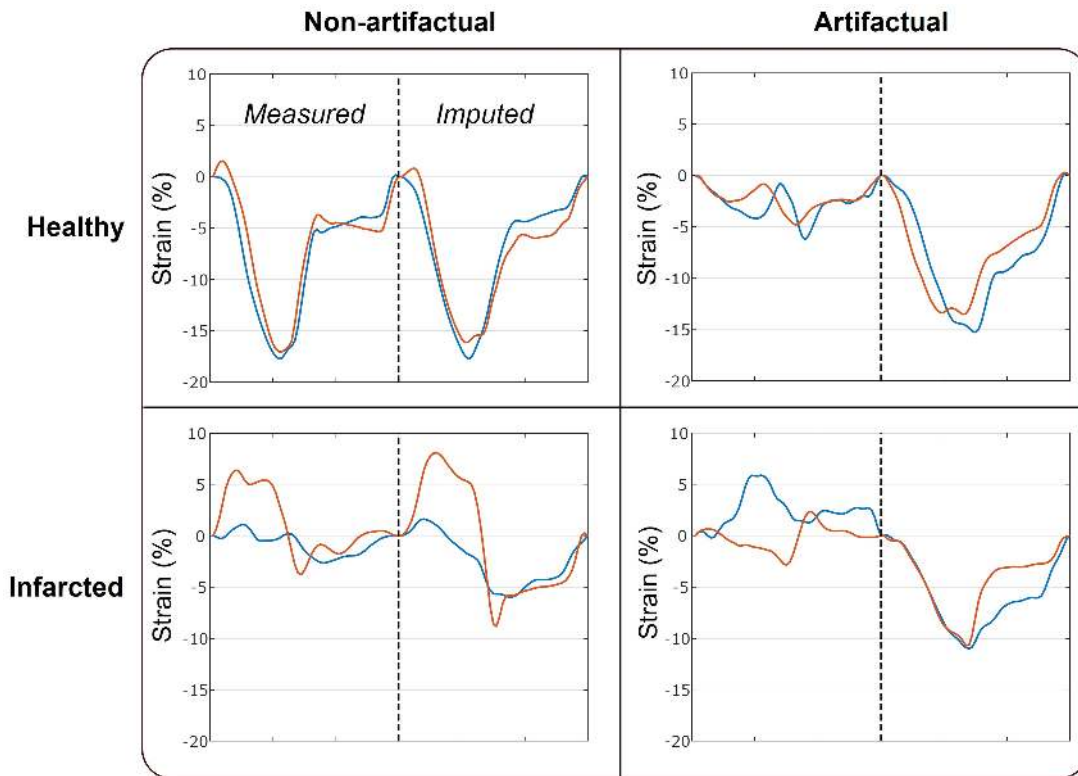
Fig. 3. Examples of the (non-)artifactual strain curves measured from healthy and infarcted myocardial segments (in the left side of each image) and their corresponding imputed traces.

tifactual strain curves by building upon a recently proposed approach for imputing missing echocardiographic strain data. Despite the practical challenges that the proposed algorithm faced, the obtained classification results confirmed the feasibility of the automatic quality assurance of the strain curves. The clinical implication of these results is significant since having feedback on the quality of the measured strain curves will facilitate their interpretation and thus increase their diagnostic value.

REFERENCES

[1] M. Leitman, P. Lysyansky, S. Sidenko *et al.*, "Two-dimensional strain-A novel software for real-time quantitative echocardiographic assessment of myocardial function," *J. Am. Soc. Echocardiogr.*, vol. 17, no. 10, pp. 1021–1029, 2004.
[2] T. Hastie, R. Tibshirani, and J. Friedman, "The elements of statistical learning," *NY Springer*, 2001.
[3] M. Tabassian, M. Alessandrini, R. Jasaityte *et al.*, "Handling missing strain (rate) curves using K-nearest neighbor imputation," in *IEEE Int. Ultrason. Symp.*, 2016, pp. 1–4.
[4] H. Khamis, A. Yahav, Z. Friedman *et al.*, "Supervised learning approach for tracking quality determination of transmural and segmental time strain curves: A feasibility study," *Journal of Biomedical Engineering and Informatics*, vol. 3, no. 2, p. 43, 2017.
[5] O. Mirea, E. D. Pagourelias, J. Duchenne *et al.*, "Intervendor differences in the accuracy of detecting regional functional abnormalities: a report from the EACVI-ASE strain standardization task force," *JACC Cardiovasc. Imaging*, vol. 11, no. 1, pp. 25–34, 2018.
[6] R. M. Lang, L. P. Badano, V. Mor-Avi *et al.*, "Recommendations for cardiac chamber quantification by echocardiography in adults: an update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging." *J. Am. Soc. Echocardiogr.*, vol. 28, no. 1, pp. 1–39.e14, 2015.
[7] O. Troyanskaya, M. Cantor, G. Sherlock *et al.*, "Missing value estimation methods for DNA microarrays," *Bioinformatics*, vol. 17, no. 6, pp. 520–525, 2001.
[8] I. Jolliffe, *Principal component analysis*. Wiley Online Library, 2002.
[9] S. A. Dudani, "The distance-weighted k-nearest-neighbor rule," *IEEE Trans. Syst. Man. Cybern.*, vol. 6, no. 4, pp. 325–327, 1976.