

ORIGINAL ARTICLES

Supervised learning approach for tracking quality determination of transmural and segmental time strain curves: A feasibility study

Hanan Khamis^{*1}, Amir Yahav¹, Zvi Friedman², Jan D'hooge³, Dan Adam¹

¹Lab for Ultrasound Signals and Image Processing and Modeling, Department of Biomedical Engineering, Technion, Haifa, Israel

²GE Ultrasound, Tirat Hacarmel, Israel

³Lab on Cardiovascular Imaging & Dynamics, Department of Cardiovascular Sciences, KU Leuven, Leuven, Belgium

Received: April 12, 2017

Accepted: June 11, 2017

Online Published: June 26, 2017

DOI: 10.5430/jbei.v3n2p43

URL: <https://doi.org/10.5430/jbei.v3n2p43>

ABSTRACT

Quantitative assessment of global and regional left ventricle function by means of myocardial strain estimation has been widely discussed as promising clinical diagnostic markers of left ventricular malfunction. These markers are provided to the clinicians without much feedback regarding their reliability, which may lead to erroneous diagnosis. Therefore, this study aims to classify the calculated strain curves into reliable or artefactual ones, before their clinical adaptation.

A supervised machine learning approach is utilized for the classification process. A total of 6,552 strain curves were used, for which a visual labeling protocol was defined and utilized by two experts.

An inter-observer labeling concordance of 93% was obtained, and a classification accuracy of 90% was achieved with a specificity of 92% and sensitivity of 78%.

This classification tool may enhance the reliability of the estimations of global, transmural and regional strain curves, by automatically classifying them into physiological or artefactual curves.

Key Words: Transmural strain imaging, Tracking quality, Time strain curves, Support vector machine, Supervised machine learning, Classification

1. INTRODUCTION

Speckle tracking echocardiography (STE) is a non-invasive technique for the assessment of left ventricle (LV) function, which mainly provides global and regional time strain curves (TSCs).^[1] Amplitude and profile analysis of these TSCs, such as global and segmental peak systolic strain, allow to detect various malfunctions of the LV myocardium.^[2-4] Currently, only peak systolic global strain is recommended as a clinical measure by the ASE/EACI.^[5] Unfortunately, there

are still some challenges that hinder the acceptance of the STE as a clinical measure. One of the main challenges relates to the semi-automatic estimation of the aforementioned clinical parameters, which produces large inter-vendor variability of the strain estimates.^[6] Once these challenges are met and consequently the STE is clinically fully accepted as part of the clinical guidelines for diagnosis and prognosis, it will probably be widely used by cardiologists as it has been recently used by researchers.

*Correspondence: Hanan Khamis; Email: khamishanan@gmail.com; Address: Lab for Ultrasound Signals and Image Processing and Modeling, Department of Biomedical Engineering, Technion, Haifa, Israel.

As part of enhancing the clinical applicability of the STE as a clinical measure, some efforts have been invested lately to detect LV abnormality by considering the entire strain temporal behavior. Tabassian et al.^[7,8] have utilized the principal component analysis as a statistical approach to classify the temporal behavior of the TSCs through the whole cardiac cycle. The rationale behind this approach is the fact that the traditional clinical landmarks (or “techno-markers”, a term used in Tabassian et al.^[7,8]) ignore the diastolic period of the cardiac cycle and evaluate the myocardial function based only on one time point. Detection results of myocardial abnormalities that were reported by Tabassian et al. have been shown to outperform the traditional methods of categorizing healthy and infarcted myocardial segments. A similar approach, reported by Clarysse et al.,^[9] has utilized statistical approaches to model the systolic phase of TSCs, extracted from tagged MRI data, instead of using the traditional landmarks.

Both approaches of myocardial strain markers and temporal behavior, together or combined, may indeed provide credible LV function analysis and diagnosis. However, erroneous strain calculation can affect the accuracy and reliability of these classification approaches due to inadequate segmentation, unreliable tracking or bad image quality, e.g. clutter, noise or out of plane motion.

Many clinical and technical reports have studied the dependence of strain and strain rate curves on image quality and the tracking algorithm. In his paper,^[10] Marwick has stated that strain and strain rate that are measured using either Doppler or STE are influenced by image quality and the measurement quality of the velocity fields. These observations have been backed by others, e.g.^[11,12] Trache et al.^[12] have shown that limited image quality has a significant impact on the agreement between 3D and 2D numerical strain values, while Hoit,^[11] has reported that measurement of strain rate (and strain) is influenced by image and signal quality, where a very high image quality is required to produce reasonable strain and strain rate results. In our previous publication,^[13] we have shown that a typical block matching tracking algorithm that uses post tracking smoothness techniques to refine the resultant strain fields, is sensitive to the characteristics and amount of smoothing, as well as to the initial segmentation of the myocardial tissue and to the noise level. Consequently, any modification that is made to the post block-matching smoothing techniques may cause the (expected) inter-vendor strain differences.^[6] We have utilized both software based phantom simulations and clinical data to evaluate the sensitivity of the calculated strain landmarks. All these works emphasize the need for a reliable measure of tracking quality.

Many other works^[2,14–19] have related to tracking quality indices that have been employed in order to decide whether to include or exclude strain measurements. The commercial vendors usually provide tracking quality indices that state a scoring of “acceptable” or “non-acceptable” segmental strain measurements, with the option of manual corrections. Generally, as mentioned by Mada et al.,^[19] tracking quality indices should indicate how well the motion of the endocardial and epicardial borders is tracked. Any software that does not provide this step may not be reliable enough and should be used in caution, as the expected strain values may be different. In their paper, Perk et al.,^[20] have listed criteria on which the tracking quality index should be based in a STE commercial software: (1) Adequate tracking is assumed when the speckle coordinates return to baseline; (2) Significant differences of tissue velocities between adjacent speckles indicate bad tracking quality, and (3) Equal strain values are expected at the beginning and at the end of heart cycle, thus a large drift indicates low tracking quality. These software based tracking quality indices can be either rejected or accepted by the user.

Unfortunately, these tracking quality indices are not always reliable, since the tracking algorithm may provide “good” tracking results even if a non-tissue object is being traced, consequently providing erroneous strain measurements. These indices also do not consider temporal behavior of each TSC (except for criteria (3) above). Additional tracking quality approach is the trashogram,^[21] which is used to identify regions of high and low strain magnitude based on the cross correlation calculated as part of the 2D phase-sensitive speckle tracking. However, despite its simplicity, trashogram may provide erroneous tracking quality estimation, since high correlation maybe a result of integer wavelength while low correlation may allow displacement estimation.^[22] To the best of our knowledge, there are no other available physiology based criteria or physiology based-tracking quality indices that may provide a direct indication whether the strain measurements or TSCs are physiological or not, i.e. clinically reliable or not. Moreover, there are no tracking quality indices that provide layer specific tracking evaluation, which allows the user to manually correct the segmentation of a specific layer (i.e. the endocardial or the epicardial layers) and/or training the software to optimize the tracking based on these quality outcomes.

In this work, we propose a new reliability measure, based on a supervised learning method, which allows to determine the reliability of transmural and segmental TSCs. Since the myocardium is roughly composed of 3 attached layers (endocardium, mid-wall and epicardium), it is logical to assume high correlation between the TSCs of these layers. Thus

high correlation between the endocardial and epicardial layers would indicate proper estimation of the physiological function, as can be understood from these studies.^[23–29] In contrast, low correlation may indicate a severe LV regional

malfunction, or bad tracking and erroneous calculations due to noise, clutter or out of plane motion, or due to incorrect and imprecise segmentation.

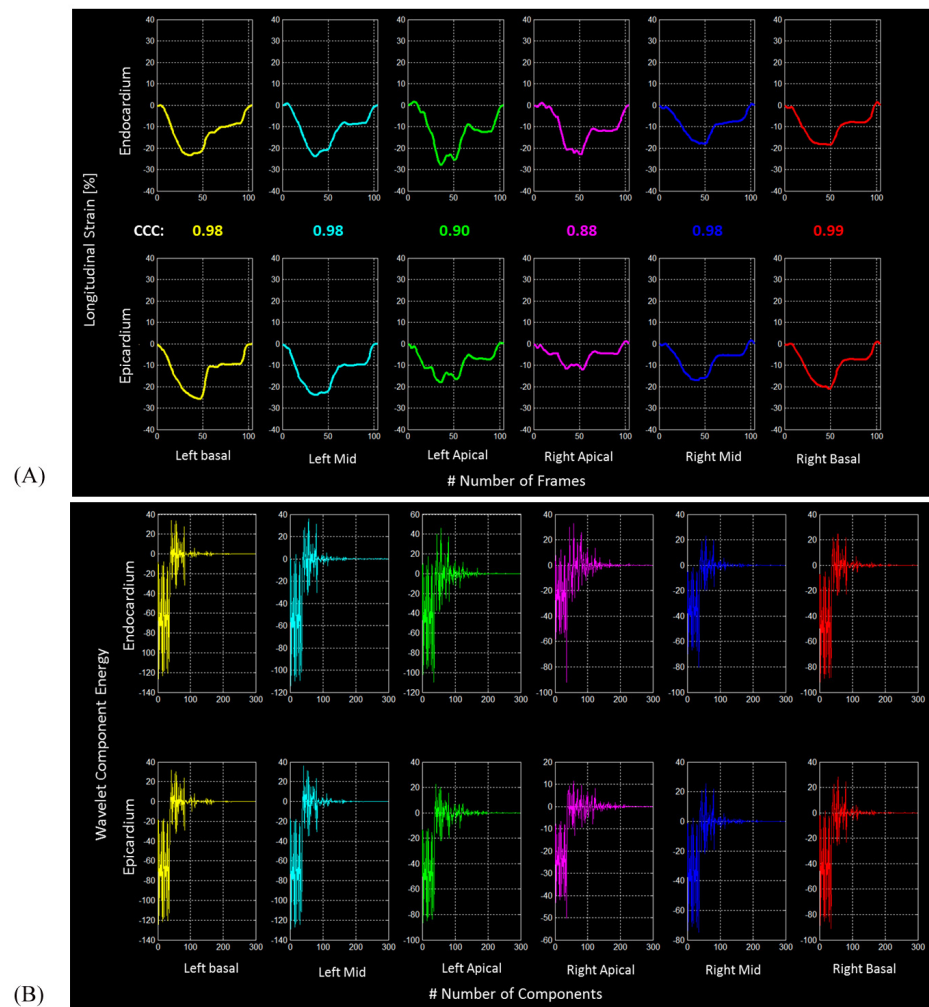


Figure 1. Example 1 of pre-labeling analysis of time-strain curve. (A) Longitudinal time-strain curves at the 6 segments of left ventricle myocardium, at the endocardial (upper row) and epicardial (bottom row) layers are shown. The cross-correlation coefficient (CCC) between the time-strain curves at the two layers is provided as well. (B) Wavelet components energy corresponding to the longitudinal time-strain curves at the 6 segments of left ventricle myocardium, at the endocardial (upper row) and epicardial (bottom row) layers are shown here. Each plot is composed of the concatenated wavelet components (left to right): Approximation at the 5th level and the Details at 5th to 1st levels.

2. MATERIALS AND METHODS

2.1 Data acquisition and preprocessing

A set of 546 echocardiograms of healthy subjects, without any known disease, from two clinical sites: Kaplan hospital in Rehovot, Israel and the University of Leipzig in Leipzig, Germany, were used in this study. The study was approved by the IRB of Kaplan Hospital, Rehovot, Israel, and the IRB of the University of Leipzig, Leipzig, Germany. Accordingly, all subjects have signed an informed consent before echocar-

diograms were acquired. The subjects' ages were between 22-82, with an average measured heart rate of 65.98 ± 9.55 beats per minute.

All echocardiograms were acquired with VIVID7 ultrasound machine (GE Vingmed Ultrasound AS, Horten, Norway) using a standard 2.5 MHz cardiac probe, according to the regular clinical protocol, including the 3 standard apical views. All subjects had a heart rate in the range of 45-90 beats per minute. Each echocardiogram was analyzed using an in-

house speckle tracking algorithm, termed “K-SAD”.[30] The myocardial walls of the LV were segmented into 6 segments following the 18-segment model.[5] Longitudinal TSCs were computed at each segment at both the endocardial and epicardial layers. A total of 6,552 TSCs were obtained. Different heart rates of the subjects yielded different number of samples in each TSC, hence, a linear interpolation was applied on all TSCs to obtain equal number of samples. No further processing was applied to the data.

2.2 Data labeling

Since the labelling of each TSC into physiological or artefactual pattern was a challenging task, a pilot study of visual qualification of 30 clips composing 540 TSCs was performed. Two experts have independently analyzed the local TSCs patterns, both in time and frequency domains, and have visualized the tracking points of the endocardial and epicardial borders superimposed on the clips’ sequences in order to understand what features may provide good indicator to identify artefactual and physiological TSCs.

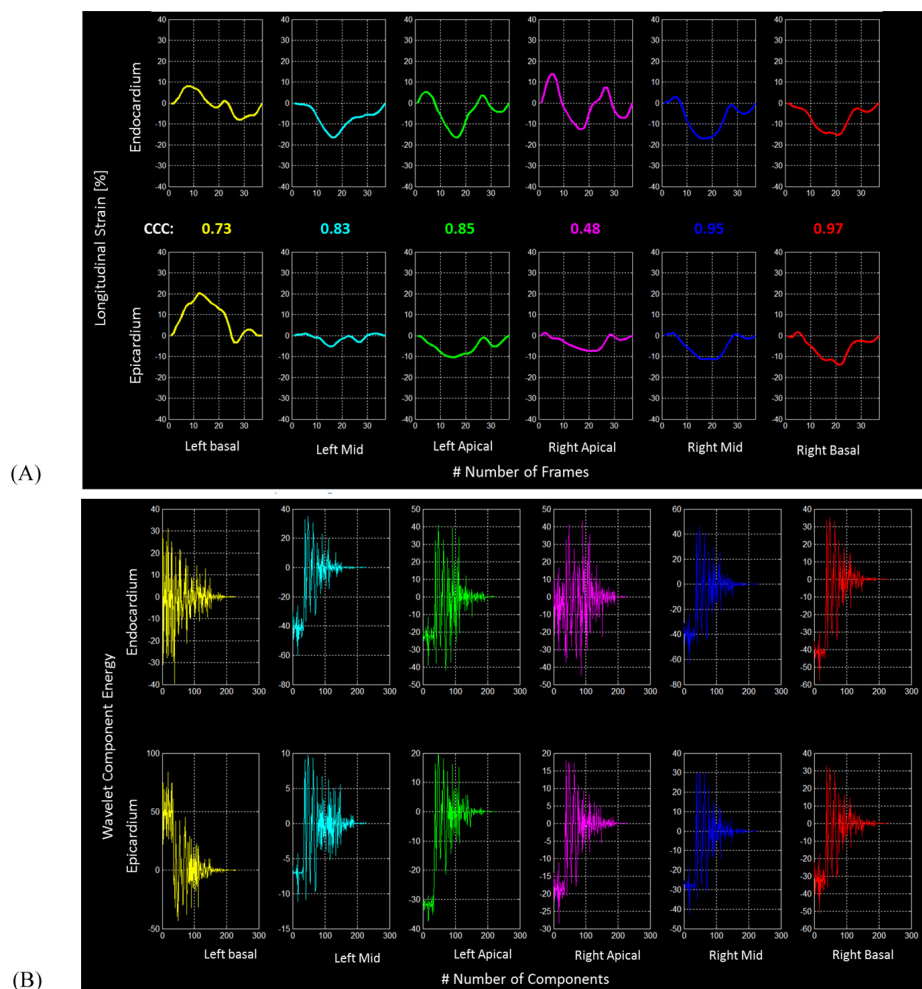


Figure 2. Example 2 of pre-labeling analysis of time-strain curve. (A) Longitudinal time-strain curves at the 6 segments of left ventricle myocardium, at the endocardial (upper row) and epicardial (bottom row) layers are shown. The cross-correlation coefficient (CCC) between the time-strain curves at the two layers is provided as well. (B) Wavelet components energy corresponding to the longitudinal time-strain curves at the 6 segments of left ventricle myocardium, at the endocardial (upper row) and epicardial (bottom row) layers are shown here. Each plot is composed of the concatenated wavelet components (left to right): Approximation at the 5th level and the Details at 5th to 1st levels.

Following this pilot study, which included physiological and clinical qualification, a list of characteristic features was defined. The two experts, in addition to their visual evalua-

tion, used these characteristic features to define the following labeling protocol for the classification process.

First, the 1D cross correlation coefficient (CCC) was calcu-

lated, between the Endocardial and Epicardial TSCs of the same segment. CCC were used to provide an initial rough prediction of whether one of the two TSCs of the same segment is artefactual. High CCC (empirically selected to be above 0.95) indicated physiological patterns at both the endocardium and the epicardium, while low CCC (empirically selected to be below 0.7) indicated that at least one of TSCs is artefactual.

Additionally, in order to exploit both time and frequency information “concealed” in each TSC, a multilevel 1D-wavelet decomposition,^[31] of each TSC was performed using Daubechies wavelet function with two vanishing moments. Shortly, multilevel 1D-wavelet decomposition employs high-pass and low-pass filters to extract the Approximation and Details coefficients which actually represent the high scale - low frequency and low scale - high frequency components of the original signal, respectively. The decomposition process is repeated N times, where N is the number of decomposition

levels. At each level, the filtering process is followed by down-sampling by two. The retrieved signals from all $N = 5$ levels (A5, D5-D1) are concatenated into one wavelet coefficient signal (WCS).

The experts performed a visual study of these WCSs, together with the rough prediction of the CCC and the behavior of the raw TSCs, to classify each TSC. The final labeling decision was visually performed by the experts, based on these three components. No automatic algorithm was performed on WCS or TSC to provide the labels. On the other hand, the classification itself was performed based on the TSCs only. In Figures 1 and 2, two examples are shown of a pre-labeling analysis of TSCs. In Figures 1A and 2A, TSCs are plotted of both the endocardium and the epicardium, at each segment. Their corresponding WCSs are plotted in Figures 1B and 2B, where the concatenated A5-D1 WCSs are shown from left to the right on the x-axis, as explained in Figure 3.

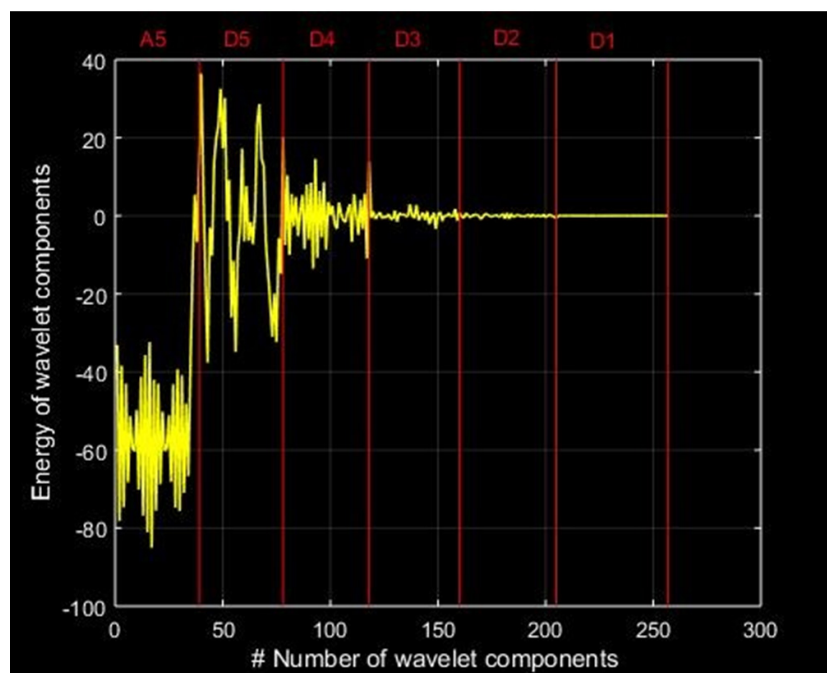


Figure 3. 1D Wavelet components signal. Wavelet components resulted from the multi-level 1D wavelet decomposition process are concatenated into one signal (from left to right): A5, D5, D4, D3, D2 and D1.

It can be seen, for example, that high CCC of 0.98-1 was calculated for the left basal and mid wall segments as well as for the right basal and mid segments (see Figure 1). This is showing the physiological profile composed of contraction of the tissue during systole (until reaching the maximum strain, in healthy subjects, at end systole) followed by relaxation of the tissue during diastole, ideally, back to its original length at end diastole.

In contrast, while looking at the apical regions of the same example (see Figure 1), one can notice a slight decrease in the CCC, and more noisy TSCs. Both curves, of the endocardium and epicardium, at the left apical segment, have two peaks. The WCSs of the TSCs at the apical segments reveal that additional energy is contained at the low Details (D3-D1), when compared to the energies contained at the same Details, at the basal and mid segments. The Details at these

levels imply high frequencies and usually represent noise. When examining the second example, presented in Figure 2, one can notice low values of CCC at the left wall (0.73, 0.83 and 0.85) and also at the right apical segment (0.48). Since CCC was a good indicator only for the extreme cases of very high or very low correlation while it was confusing for most of the cases in between, the information revealed from the TSC and WCS were more powerful for the visual labeling process. In the same example, the TSC of the left basal behaves quite differently from any expected physiological profile, of either the endocardium or the epicardium.

Double positive peaks appear in the TSC of the endocardial layer, of both apical regions. In addition, one can notice untypical profiles of the TSCs of the epicardium of the left mid, apical and right apical segments. The visualization of the WCSs' energies added better insight, by noting the location and intensity of the energy of each component. At the left basal segment, the energy of the Approximation (A5) at the endocardium is centered around zero, which is not the case for physiological TSCs. At the epicardium, the energy is centered around a positive energy, indicating expansion of the tissue instead of contraction.

Table 1. Visual labeling protocol

	Physiology	Non-physiology
CCC	> 0.95	< 0.70
TSC	Composed of left ventricle contraction phases: Contraction during systole, relaxation during diastole One negative peak	Appearance of several peaks Positive contraction during systole Noisy Fluctuations around zero
WCS	A5 energy is centered at a negative value D5-D1 are centered at zero Energy of D4 is significantly lower than D5	A5 and D5 have similar energies Energy of A5 is centered at zero Noisy D3-D1

Note. Table 1 presents the rules for physiological and non-physiological curves that were defined for the cross correlation coefficient (CCC), time strain curves (TSC) and wavelet components signals (WCS) based on the training set.

Table 1 lists the different visual behavioral features that were roughly utilized in the visual labeling process. The grading protocol was performed as follows: physiological TSCs were labeled by 1, while artefactual TSCs were labeled by 0. The remaining TSCs, which the experts had difficulty in labeling, were considered as “undetermined TSCs”, and were labeled by either 2, for TSCs that were more likely to be physiological than artefactual, or by 3 for TSCs that were more likely artefactual than physiological (see Results sections).

2.3 Support vector machine

Support vector machine (SVM) is a statistical learning theory approach, first introduced by Vapnik and Boser et al.,^[32,33] and was successively developed by different researchers, among them are Cristianini and Shawe-Taylor.^[34] The robustness and high performance of the SVMs, regardless of noise and complexity of the data, made them widely used in multiple applications such as pattern recognition and supervised classification problems.^[35-37]

For classification purposes, SVMs are trained to discriminate between two or more classes, using a labeled data set: given a training data set, SVM defines an optimal hyperplane for the separation between different classes within the training set, such that each sample in one class has the largest distance (maximal margin) from other samples of different class. This

hyperplane is called the “maximal margin hyperplane” and the training set used is called “support vectors”.

Since linear separation usually doesn't fit the data sets that need to be discriminated, a Kernel method based SVM is utilized: Let $D = \{(x_i, y_i) \mid x_i \in R^p, y_i \in \{0, 1\}_{i=1}^n\}$ be a training set of n sample vectors. x_i is a p -dimensional vector labeled by y_i , which indicates the class it belongs to.

A general SVM classifier is obtained by solving the convex Lagrange dual of the primal maximal margin, as follows:

$$f(x) = \sum_{i=1}^n \alpha_i \cdot y_i \cdot K(x, x_i) + b \quad (1)$$

where $K(x, x_i)$ is the kernel function, x is the data point and α_i are the Lagrange multipliers that define the combination of the support vectors. Usually, only a subset of α_i are of non-zero values, thus defining the supports vectors that lie closest to the discrimination hyperplane. The two most widely used kernels are the “polynomial kernels” and the “Gaussian radial basis function kernels”. A Gaussian kernel function is given as:

$$K(x, x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (2)$$

where σ is the scaling factor of the radial Gaussian basis

function, which defines the width (standard deviation) of the Gaussian function. A Gaussian kernel that is computed with support vectors is an exponentially decaying function in the input feature space, while it decays uniformly in all directions as one recedes from the feature space center. A Gaussian based SVM classifier can be simply explained as a weighted linear combination of the kernel function, computed between a data point and each of the support vectors.

When noise is present, a simplified maximum margin algorithm may fail due to overfitting, while attempting to find an optimal discrimination that does not allow tolerance in the training error. Consequently, a widely used modification employs some regularity constraints, in order to trade some training accuracy for a more powerful prediction. These constraints are named the Karush-Kuhn-Tucker (KKT) conditions after Harold W. Kuhn, Albert W. Tucker and William Karush.^[38,39]

Implementation: For the classification of the TSCs, a set of SVM functions were utilized for training and testing, which are available as part of the Statistics and Machine Learning Toolbox™ (MATLAB R2015b, The Mathworks Inc. Natick, MA, 2015). A Gaussian kernel, RBF, was used as a kernel function, with a scaling factor σ of different values, for the purpose of performance and sensitivity testing. A Sequential Minimal Optimization based method was used as well, with the default KKT value, for defining the discriminative hyperplane.

2.4 Classification

As part of the classification process, the interpolated and labeled TSCs were divided into training and test sets (to include 70% and 30%, respectively, of the TSCs). The kernel-based SVM was trained using the labeled training set, to provide a set of representative vectors and define a hyperplane for the classification. The SVMTRAIN MATLAB function was used for that. Then, the trained classifier was verified using the test set and the SVMCLASSIFY MATLAB function. Both SVMTRAIN and SVMCLASSIFY MATLAB functions were available using MATLAB R2015b (The Mathworks Inc. Natick, MA, 2015). The classification process, including the training stage, was performed three times independently, using different, but overlapping data sets:

Data set 1: Only TSCs that were successfully labeled by 1 or 0 were included, while all undetermined TSCs (labeled by 2 or 3) were excluded.

Data set 2: All TSCs were included, while the undetermined TSCs (labeled by 2 or 3) were all roughly considered as artefactual TSCs, and labeled by 0.

Data set 3: All TSCs were included, while the undetermined

TSCs that were “more likely to be physiologic” were relabeled by 1 (physiology), while those that were “more likely to be artefactual” were relabeled by 0 (artefactual).

The accuracy, sensitivity and specificity of the classification were calculated for the test set of each data set. The accuracy is calculated as:

$$\frac{TP + TN}{TP + TP + TN + FN} \quad (3)$$

where TP is “true positive” which indicates the number of physiological TSCs that were classified as physiological, TN is “true negative” which indicates the number of artefactual TSCs that were classified as artefactual, FP is “false positive” which indicates the number of artefactual TSCs that were classified as physiological and FN is “false negative” which indicates the number of physiological TSCs that were classified as artefactual. Consequently, the sensitivity measures the proportion of physiological TSCs that were correctly classified as such, i.e. $\frac{TP}{(TP+FN)}$ and the specificity measures the proportion of artefactual TSCs that were correctly classified as such, i.e. $\frac{TN}{(TN+FP)}$.

The classification results were then used to determine the tracking quality of each segment along the myocardium, and to indicate the reliability of each TSC prior to the clinical diagnosis. In addition, since the classification was dependent on the scaling factor σ , of the Gaussian kernel based SVM, the accuracy has been evaluated also as a function of σ , for each data set.

Table 2. Groups of TSCs

	#Total	#Training set		#Test set	
		4000		1715	
Data set 1	5715	3690	310	1405	310
		'1'	'0'	'1'	'0'
		4586		1966	
Data set 2	6552	3857	729	1238	728
		'1'	'0'	'1'	'0'
		4586		1966	
Data set 3	6552	4164	422	1545	421
		'1'	'0'	'1'	'0'

Note. Table 2 summarizes the number of time strain curves used for the training and test sets, in the three data sets: Data set 1, Data set 2 and Data set 3.

3. RESULTS

3.1 Time-strain curve labeling

Out of all the 6,552 TSCs, 837 TSCs were found to be challenging during the labeling process, and were labeled as the undetermined group. Among them, 614 TSCs were labeled by “2” (more likely to be physiological) and 223 TSCs were labeled by “3” (more likely to be artefactual). The number

of TSCs that were used for training and testing, following the three data sets defined above, are listed in Table 2. The classification concordance between the two experts was 93%. The rest 7% were labeled as “more likely to be artefactual” (i.e. “3”).

It is interesting to note that out of all TSCs that were labeled as artefactual curves, 57.74% were measured at the subepicar-

dial layer, and 42.26% were measured at the subendocardial layer. Out of all segments that were labeled as artefactual (“0”) or “more likely to be artefactual” (“3”), 66.31% were measured at the apical segments, while only 10.56% and 23.13% of these were measured at the mid-wall and basal segments, respectively. Additionally, 52.81% of all TSCs that were labeled as undetermined (with “2” or “3”), were measured at the epicardial layer.

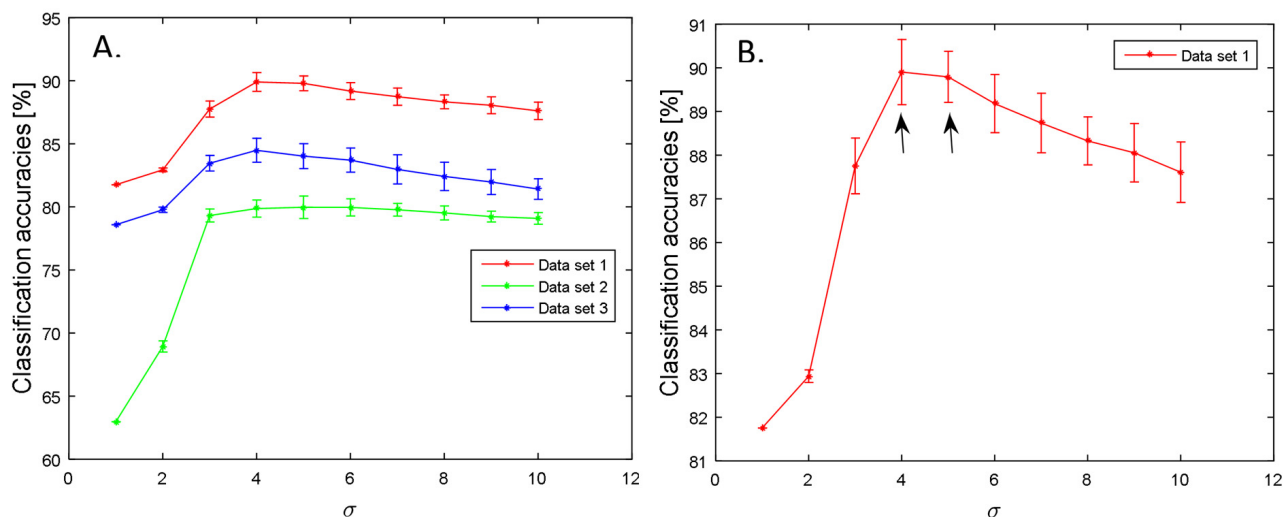


Figure 4. TSC classification accuracies. (A) TSC classification accuracies as a function of the scaling factor, σ , of the Gaussian kernel used in the SVM classifier for Data set 1,2 and 3. (B) TSC classification accuracies as a function of the scaling factor, σ , of the Gaussian kernel used in the SVM classifier for Data set 1 only (a zoomed graph).

Table 3. The classification results and confusion matrix

A. Classification results

	Accuracy	Sensitivity	Specificity
Data set 1	90%	92%	78%
Data set 2	80%	84%	73%
Data set 3	85%	90%	85%

B. Confusion matrix

Prediction\ true label	Physiologic curve	Artefactual curve
Data set 1		
Physiologic curve	1327 (\cong 94.4%)	78
Artefactual curve	85	225 (\cong 73%)
Data set 2		
Physiologic curve	1076 (\cong 87%)	162
Artefactual curve	234	494 (\cong 68%)
Data set 3		
Physiologic curve	1394 (\cong 90%)	151
Artefactual curve	154	267 (\cong 63%)

Note. Table 3 (A) presents the classification accuracy, sensitivity and specificity yielded by the three data sets, while (B) presents the confusion matrix for both the physiologic and artefactual curves, per each data set.

3.2 Classification accuracy

3.2.1 Sensitivity to scaling factor, σ :

The classification accuracy reached its maximal value, for each one of the data sets, when σ was in the range of [4-5], as depicted in Figure 4A and Figure 4B. For all data sets, lower values of σ provided lower accuracies, while higher values of σ showed a decrease in the accuracy rate.

Data set 1, which did not include the undetermined traces, have yielded the highest accuracies for all values of σ , in comparison to data set 2 and 3. Classification accuracies for all values of σ were the lowest (see Figure 4A).

3.2.2 Accuracy, sensitivity and specificity analysis

Using the scaling factors $\sigma_1 = 4$, $\sigma_2 = 5$ and $\sigma_3 = 4$, the best classification accuracy rates were obtained for Data set 1, 2 and 3, respectively. The highest accuracy rate, 90%, was achieved for Data set 1 with sensitivity of 92% and specificity of 78%. Data set 3 yielded an accuracy of 85% with sensitivity of 90% and specificity of 85%, while data set 2 yielded an accuracy of 80% with sensitivity and specificity rates of 84% and 73%, respectively. The results are reported in Table 3A. In addition, confusion matrix per each data set is presented in Table 3B.

4. DISCUSSION

The highest classification accuracy has been provided by data set 1, as expected, where all undetermined TSCs were excluded. The inclusion of these curves makes the training more challenging, and thus may affect the classification process.

The case of considering all undetermined TSCs as artefactual curves, as in data set 2, may be acceptable for clinical applications, since it is advisable to reject such curves instead of taking the risk of providing an erroneous clinical classification (thus decision). However, for the training stage, it obviously increases the classification accuracies that might be caused by incorrect labeling. Equivalently, splitting the undetermined TSCs into “more likely to be physiologic” and “more likely to be artefactual”, was sufficient to decrease the number of incorrect labeling and increase the classification accuracies.

Sensitivity was found to be higher for data set 1, as expected. The specificity was found to be slightly higher for data set 3 than data set 1, due to the variability in the artefactual TSCs. As explained earlier, a physiological normal curve has a unique behavior, while artefactual curves may obtain many different temporal behaviors. Since “specificity” measures the proportion of the artefactual TSCs that were correctly identified as artefactual, the additional curves that were added to the artefactual training set of TSCs have expanded

the ability to represent more options of artefactual behavior, allowing better identification of such cases. The Confusion matrix results show that the accuracy for the physiological curves was better than that of the artefactual curves. This could be explained by the heterogeneity of the artefactual curves. There is no recurrent pattern or behavior among the artefactual curves, which may lead to confusion in the classification itself. Another potential explanation to these results is the unequal ratio between the two classes among the training and test stages. Unequal ratios is one of the limitations in this study, since it may lead to biased results. One way of targeting these issues is by enlarging the artefactual data set. Despite this limitation, the results are satisfying with more than 70% of accuracy for the artefactual curves and more than 94% of accuracy for the physiological curves.

Finally, one may notice that the results of the visual labelling revealed that most measurements of low tracking quality were obtained at the apical regions. This is due to inaccurate detection of the myocardial borders, near field noise, off-axis acquisitions and out-of-plane motion. These constraints, either all of them together, or apart, may severely influence the strain calculations.

We believe that the approaches presented by Tabassian et al.^[7,8] and Clarysse et al.^[9] are promising, as they provide more reliable TSC analysis and clinical diagnosis than the traditional approaches, which are extensively used by most technical papers analyzing performance comparison studies^[2,40-43] as well as by many clinical studies.^[44-46] Yet, we also believe that a reliability determination of these results prior to any clinical diagnosis is essential.

Since the emergence of strain-based techniques for the assessment of LV function and until recently, only landmarks of “one time point” strain (or strain rate) have been measured and commonly used by many of the commercial software products for LV functional analysis. These commercial software packages are based on different techniques such as Speckle tracking, cine MRI deformable registration and MR feature tracking. Performance evaluation of these software packages is based mainly on the “one time point” global and segmental strain values, e.g.^[40,42,43] Similarly, multicenter studies, such as Marwick et al.,^[2] and meta-analysis reports, such as,^[41,46] which usually influence the clinical guidelines and define the normal and abnormal ranges of parameters, as well as clinical decisions, are also based on these “one time point” landmarks. In addition, many other clinical publications,^[44,45] have investigated the potential of using such clinical strain landmarks, in addition to or instead of the conventional clinical parameters (such as ejection fraction, end systolic and end diastolic left ventricle areas/volumes, etc.).

Therefore, whether the whole TSCs or only “one time point” landmarks are used for the clinical diagnosis, it is crucial to perform a reliability determination of the obtained TSCs in order to exclude artefactual results and provide a more accurate and reliable clinical diagnosis. Finally, an extension to this promising technique to detect artefactual patterns in pathological TSCs will indeed push the STE approach forward toward its full acceptance in the daily clinical protocols.

We have utilized for the classification process the non-linear SVM algorithm, which is implemented in MATLAB as part of the machine learning toolbox. One could consider for this classification alternative supervised classifiers (e.g. those others available in the MATLAB toolbox), such as K-Nearest Neighbor (KNN) or Discriminant Analysis (DA). However, one must be aware that although KNN is also a non-linear classifier, it needs to be carefully tuned by choosing an appropriate K value and distance metric. If DA is considered, one should use non-linear versions of it, e.g. using a kernel, as for the SVM. More advanced deep learning techniques could also be applied, such as neural network. For the current feasibility study, the SVM was chosen since it is well known as a robust classifier.^[35–37,47]

In this work, we have utilized only our in-house K-SAD STE algorithm, since it was shown to be more accurate than the most popular commercial product.^[30] This limitation is not significant since the main attempt here was to demonstrate the reliability of the SVM classification on some STE algorithm. In addition, we have analyzed normal subjects only, thus the classification into physiological and artefactual curves is relatively simple. This is, however, a mandatory first step, by itself important, before relating TSCs detected as artefactual to local or global pathology or abnormality. It is hypothesized that for the clinical pathologies that cause global deterioration of the LV function, e.g. cardiotoxicity due to chemotherapy treatment, the damage at the early stages would cause global impairment of the endocardial layer.^[48] Under such conditions, the cross correlation between TSCs at the different myocardial layers should remain high, since only the amplitude may decrease, while the TSC profile would not change. This may be true also for myocardial remodeling caused by aortic stenosis, or increased afterload. As for other pathological cases, such as transmural myocardial infarction (MI), which affects essentially the full thickness of the wall but is localized to one or two segments, the cross correlation may be decreased, as the patterns of both layers may be distorted/defaced, e.g. significant changes of the time to post systolic shortening.^[49] This is one of the major limitations of the proposed method, since in some

cases, a combination of bad image quality and a localized myocardial pathology may be challenging. Extending and generalizing the method presented here would require significant efforts, so that it would be able to determine whether a given TSC pattern is abnormal or artefactual. Yet, the current version of the method allows one to examine the reliability of the results, and underline those segments that are problematic – either artefactual or pathological, allowing the clinician to make the final decision. Also, the current version of the method provides a feedback to the tracking algorithm, whether to perform another iteration so as to produce more accurate results.

5. CONCLUSIONS

In this work, we have shown that for those several cases, when the transmural and segmental TSCs may be clinically unreliable, due to inadequate myocardial segmentation or deficient tracking performance, it is essential to automatically detect and classify the artefactual TSCs before clinical analysis, thus potentially reducing erroneous diagnosis. Supervised learning based SVM model was employed here for the classification task. The results indicate that most of the artefactual TSCs were calculated at the epicardium, which apparently emphasizes the challenges of epicardial segmentation. Artefactual or unreliable results may severely affect the whole analysis process, regardless of whether the calculations are of transmural or non-transmural, segmental or global strains. Once a TSC, calculated at a specific segment and/or myocardial layer, is detected and classified as an artefactual, it may be used for modifying and re-performing the whole calculation process: modifying the segmentation or tracking calculations, and/or selecting a new clip or part of a clip with reduced presence of clutter, noise or out-of-plane motion. Consequently, the strain analysis may be repeated or rejected at the specific segment or layer. It is also perceived that with the recently developed ultrasound systems that feature novel beamforming techniques with better spatial resolution and less clutter noise, specifically at the apical regions, it will be hopefully possible to increase the performance of the tracking algorithms. This will allow acquisition of a much larger percentage of physiological TSCs, and significantly smaller percentage of artefactual TSCs.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Sarah Shimoni from Kaplan Hospital and Dr. Andreas Hagedorff from Leipzig University for their assistance in data acquisition. This study was approved by the IRB of Kaplan Hospital, Rehovot, Israel, and the IRB of the University of Leipzig, Germany.

REFERENCES

- [1] D'hooge J, Bijmens B, Thoen J, et al. Echocardiographic strain and strain-rate imaging: a new tool to study regional myocardial function. *Medical Imaging*. IEEE Transactions on Medical Imaging. 2002; 21(9): 1022-30. PMID:12564870. <https://doi.org/10.1109/TMI.2002.804440>
- [2] Marwick TH, Leano RL, Brown J, et al. Myocardial strain measurement with 2-dimensional speckle-tracking echocardiography: definition of normal range. *JACC: Cardiovascular Imaging*. 2009; 2(1): 80-4. PMID:19356538. <https://doi.org/10.1016/j.jcmg.2007.12.007>
- [3] Armstrong G, Pasquet A, Fukamachi K, et al. Use of peak systolic strain as an index of regional left ventricular function: comparison with tissue Doppler velocity during dobutamine stress and myocardial ischemia. *Journal of the American Society of Echocardiography*. 2000; 13(8): 731-7. PMID:10936816. <https://doi.org/10.1067/mje.2000.105912>
- [4] Popović ZB, Benejam C, Bian J, et al. Speckle-tracking echocardiography correctly identifies segmental left ventricular dysfunction induced by scarring in a rat model of myocardial infarction. *American Journal of Physiology-Heart and Circulatory Physiology*. 2007; 292(6): H2809-16. PMID:17277023. <https://doi.org/10.1152/ajpheart.01176.2006>
- [5] Lang RM, Badano LP, Mor-Avi V, 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. *Journal of the American Society of Echocardiography*. 2015; 28(1): 1-39. PMID:25559473. <https://doi.org/10.1016/j.echo.2014.10.003>
- [6] Farsalinos KE, Daraban AM, Ünü S, et al. Head-to-head comparison of global longitudinal strain measurements among nine different vendors: The EACVI/ASE inter-vendor comparison study. *Journal of the American Society of Echocardiography*. 2015; 28(10): 1171-81. PMID:26209911. <https://doi.org/10.1016/j.echo.2015.06.011>
- [7] Tabassian M, Alessandrini M, De Marchi L, et al. Principal Component Analysis for the Classification of Cardiac Motion Abnormalities Based on Echocardiographic Strain and Strain Rate Imaging. In *Functional Imaging and Modeling of the Heart*, Springer International Publishing. 2015; 9126: 83-90. https://doi.org/10.1007/978-3-319-20309-6_10
- [8] Tabassian M, Alessandrini M, Herbots L, et al. Automatic detection of ischemic myocardium by spatio-temporal analysis of echocardiographic strain and strain rate curves. 2015 IEEE International Ultrasonics Symposium (IUS). 2015: 1-4. <https://doi.org/10.1109/ULTSYM.2015.0107>
- [9] Clarysse P, Han M, Croisille P, et al. Exploratory analysis of the spatio-temporal deformation of the myocardium during systole from tagged MRI. *IEEE Transactions on Biomedical Engineering*. 2002; 49(11): 1328-39. PMID:12450363. <https://doi.org/10.1109/TBME.2002.804587>
- [10] Marwick TH. Measurement of strain and strain rate by echocardiography: ready for prime time? *Journal of the American College of Cardiology*. 2006; 47(7): 1313-27. PMID:16580516. <https://doi.org/10.1016/j.jacc.2005.11.063>
- [11] Hoit BD. Strain and strain rate echocardiography and coronary artery disease. *Circulation: Cardiovascular Imaging*. 2011; 4(2): 179-90. PMID:21406664. <https://doi.org/10.1161/CIRCIMAGING.110.959817>
- [12] Trache T, Stöbe S, Tarr A, et al. The agreement between 3D, standard 2D and triplane 2D speckle tracking: effects of image quality and 3D volume rate. *Echo Research and Practice*. 2014; 1(2): 71-83.
- [13] Khamis H, Shimoni S, Hagendorff A, et al. Feasibility of reproducible vendor independent estimation of cardiac function based on first generation speckle tracking echocardiography. *Journal of Biomedical Engineering and Informatics*. 2015; 2(2): 57. <https://doi.org/10.5430/jbei.v2n2p57>
- [14] Cameli M, Caputo M, Mondillo S, et al. Feasibility and reference values of left atrial longitudinal strain imaging by two-dimensional speckle tracking. *Cardiovascular Ultrasound*. 2009; 7(1): 6. PMID:19200402. <https://doi.org/10.1186/1476-7120-7-6>
- [15] Di Salvo G, Russo MG, Paladini D, et al. Two-dimensional strain to assess regional left and right ventricular longitudinal function in 100 normal foetuses. *European Heart Journal-Cardiovascular Imaging*. 2008; 9(6): 754-6. <https://doi.org/10.1093/ejehocard/jen134>
- [16] Brown J, Jenkins C, Marwick TH. Use of myocardial strain to assess global left ventricular function: A comparison with cardiac magnetic resonance and 3-dimensional echocardiography. *American Heart Journal*. 2009; 157(1): 102.e1-5. PMID:19081404. <https://doi.org/10.1016/j.ahj.2008.08.032>
- [17] Belghiti H, Brette S, Lafitte S, et al. Automated function imaging: a new operator-independent strain method for assessing left ventricular function. *Archives of Cardiovascular Diseases*. 2008; 101(3): 163-9.
- [18] Becker M, Bilke E, Kühl H, et al. Analysis of myocardial deformation based on pixel tracking in two dimensional echocardiographic images enables quantitative assessment of regional left ventricular function. *Heart*. 2006; 92(8): 1102-8. PMID:16387826. <https://doi.org/10.1136/hrt.2005.077107>
- [19] Mada RO, Duchenne J, Voigt JU. Tissue Doppler, Strain and Strain Rate in ischemic heart disease "How I do it". *Cardiovascular Ultrasound*. 2014; 12(1): 1. PMID:25234056. <https://doi.org/10.1186/1476-7120-12-38>
- [20] Perk G, Tunick PA, Kronzon I. Non-Doppler two-dimensional strain imaging by echocardiography—from technical considerations to clinical applications. *Journal of the American Society of Echocardiography*. 2007; 20(3): 234-43. PMID:17336748. <https://doi.org/10.1016/j.echo.2006.08.023>
- [21] Kaluzynski K, Chen X, Emelianov SY, et al. Strain rate imaging using two-dimensional speckle tracking. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*. 2001; 48(4): 1111-23. PMID:11477770. <https://doi.org/10.1109/58.935730>
- [22] Hall TJ, Jiang J. Motion tracking for palpation imaging. *Biomedical Imaging: Nano to Macro*. 2004. IEEE International Symposium on. IEEE. 2004: 45-8. <https://doi.org/10.1109/isbi.2004.1398470>
- [23] Nguyen TT, Espinoza AW, Hyler S, et al. Transmural strain distribution across the cardiac wall and its dependency on measurement site. 2012 IEEE International Ultrasonics Symposium (IUS). 2012: 1-4. <https://doi.org/10.1109/ultsym.2012.0271>
- [24] Leitman M, Lysiansky M, Lysiansky P, et al. Circumferential and longitudinal strain in 3 myocardial layers in normal subjects and in patients with regional left ventricular dysfunction. *Journal of the American Society of Echocardiography*. 2010; 23(1): 64-70. PMID:20122496. <https://doi.org/10.1016/j.echo.2009.10.004>
- [25] Adamu U, Schmitz F, Becker M, et al. Advanced speckle tracking echocardiography allowing a three-myocardial layer-specific analysis of deformation parameters. *European Heart Journal-Cardiovascular Imaging*. 2008; 10(2): 303-8. <https://doi.org/10.1093/ejehocard/jen238>
- [26] Ledesma-Carbayo MJ, Kybic J, Desco M, et al. Cardiac motion analysis from ultrasound sequences using non-rigid registration. In-

- ternational Conference on Medical Image Computing and Computer-Assisted Intervention, Springer Berlin Heidelberg. 2001; 2208: 889-96. https://doi.org/10.1007/3-540-45468-3_106
- [27] Ledesma-Carbayo MJ, Kybic J, Suhling M, et al. Cardiac ultrasound motion detection by elastic registration exploiting temporal coherence. *Biomedical Imaging, 2002. Proceedings. 2002 IEEE International Symposium on IEEE*. 2002: 585-8. <https://doi.org/10.1109/isbi.2002.1029325>
- [28] Sakurai D, Asanuma T, Masuda K, et al. Myocardial layer-specific analysis of ischemic memory using speckle tracking echocardiography. *The International Journal of Cardiovascular Imaging*. 2014; 30(4): 739-48. PMID:24532055. <https://doi.org/10.1007/s10554-014-0388-x>
- [29] Shi J, Pan C, Kong D, et al. Left Ventricular Longitudinal and Circumferential Layer-Specific Myocardial Strains and Their Determinants in Healthy Subjects. *Echocardiography*. 2015; 33(4). PMID:26661049.
- [30] Khamis H, Shimoni S, Hagendorff A, et al. Optimization Based Speckle Tracking Algorithm for Left Ventricle Strain Estimation: A Feasibility Study. *IEEE Transactions on Ultrasonics Ferroelectrics, and Frequency Control*. 2016; 63(8): 1093-106. PMID:27214894. <https://doi.org/10.1109/TUFFC.2016.2569619>
- [31] Mallat SG. A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1989; 11(7): 674-93. <https://doi.org/10.1109/34.192463>
- [32] Vapnik VN. *Statistical learning theory*. New York: Wiley. 1998: 1.
- [33] Boser BE, Guyon IM, Vapnik VN. A training algorithm for optimal margin classifiers. *Proceedings of the fifth annual workshop on Computational learning theory*. ACM. 1992: 144-52. <https://doi.org/10.1145/130385.130401>
- [34] Cristianini N, Shawe-Taylor J. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press. 2000; 22(2). <https://doi.org/10.1017/CB09780511801389>
- [35] Burges CJ. A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*. 1998; 2(2): 121-67. <https://doi.org/10.1023/A:1009715923555>
- [36] Hsu CW, Chang CC, Lin CJ. *A practical guide to support vector classification*. 2003: 1-16.
- [37] Fradkin D, Muchnik I. Support vector machines for classification. *Discrete Methods in Epidemiology*. 2006; 70: 13-20.
- [38] Kuhn HW. *Nonlinear Programming: A Historical View Traces and Emergence of Nonlinear Programming*. Springer Basel. 2014: 393-414. https://doi.org/10.1007/978-3-0348-0439-4_18
- [39] Kjeldsen TH. A contextualized historical analysis of the Kuhn-Tucker Theorem in nonlinear programming: The impact of World War II. *Historia Mathematica*. 2000; 27(4): 331-61. <https://doi.org/10.1006/hmat.2000.2289>
- [40] Lamacie MM, Thavendiranathan P, Hanneman K, et al. Quantification of global myocardial function by cine MRI deformable registration-based analysis: Comparison with MR feature tracking and speckle-tracking echocardiography. *European Radiology*. 2016: 1-12.
- [41] Levy PT, Macheffsky A, Sanchez AA, et al. Reference ranges of left ventricular strain measures by two-dimensional speckle-tracking echocardiography in children: a systematic review and meta-analysis. *Journal of the American Society of Echocardiography*. 2016; 29(3): 209-25. PMID:26747685. <https://doi.org/10.1016/j.echo.2015.11.016>
- [42] Ledesma-Carbayo MJ, Santos A, Kybic J, et al. Myocardial strain analysis of echocardiographic sequences using nonrigid registration. *Computers in Cardiology*. 2004: 313-6.
- [43] Evin M, Cluzel P, Lamy J, et al. Assessment of left atrial function by MRI myocardial feature tracking. *Journal of Magnetic Resonance Imaging*. 2015; 42(2): 379-89. PMID:25630749. <https://doi.org/10.1002/jmri.24851>
- [44] Barbier P, Mirea O, Cefalù C, et al. Reliability and feasibility of longitudinal AFI global and segmental strain compared with 2D left ventricular volumes and ejection fraction: intra- and inter-operator, test-retest, and inter-cycle reproducibility. *Eur Heart J Cardiovasc Imaging*. 2015; 16(6): 642-52. PMID:25564395. <https://doi.org/10.1093/ehjci/jeu274>
- [45] Knackstedt C, Bekkers SC, Schummers G, et al. Fully automated versus standard tracking of left ventricular ejection fraction and longitudinal strain: the FAST-EFs Multicenter Study. *Journal of the American College of Cardiology*. 2015; 66(13): 1456-66. PMID:26403342. <https://doi.org/10.1016/j.jacc.2015.07.052>
- [46] Liou K, Negishi K, Ho S, et al. Detection of Obstructive Coronary Artery Disease Using Peak Systolic Global Longitudinal Strain Derived by Two-Dimensional Speckle-Tracking: A Systematic Review and Meta-Analysis. *Journal of the American Society of Echocardiography*. 2016; 29(8): 724-35. PMID:27155815. <https://doi.org/10.1016/j.echo.2016.03.002>
- [47] Byvatov E, Fechner U, Sadowski J, et al. Comparison of support vector machine and artificial neural network systems for drug/non-drug classification. *Journal of Chemical Information and Computer Sciences*. 2003; 43(6): 1882-9. PMID:14632437. <https://doi.org/10.1021/ci0341161>
- [48] Thavendiranathan P, Poulin F, Lim KD, et al. Use of myocardial strain imaging by echocardiography for the early detection of cardiotoxicity in patients during and after cancer chemotherapy: A systematic review. *Journal of the American College of Cardiology*. 2014; 63(25_PA): 2751-68.
- [49] Gjesdal O, Hopp E, Vartdal T, et al. Global longitudinal strain measured by two-dimensional speckle tracking echocardiography is closely related to myocardial infarct size in chronic ischaemic heart disease. *Clinical Science*. 2007; 113(6): 287-96. PMID:17501720. <https://doi.org/10.1042/CS20070066>