Published in final edited form as:

Ann Biomed Eng. 2021 February ; 49(2): 922–932. doi:10.1007/s10439-020-02639-1.

Machine Learning Outcome Prediction in Dilated Cardiomyopathy Using Regional Left Ventricular Multiparametric Strain

Robert M. MacGregor, MD¹, Aixia Guo, PhD², Muhammad F. Masood, MD¹, Brian P. Cupps, PhD¹, Gregory A. Ewald, MD³, Michael K. Pasque, MD^{1,*}, Randi Foraker, PhD²

¹Department of Surgery, Division of Cardiothoracic Surgery, Washington University School of Medicine, Barnes-Jewish Hospital, St. Louis, Missouri

²Institute for Informatics, Division of General Medical Sciences, Washington University School of Medicine, St. Louis, Missouri

³John T. Milliken Department of Internal Medicine, Cardiovascular Division, Washington University School of Medicine, St. Louis, Missouri

Abstract

Background: The clinical presentation of idiopathic dilated cardiomyopathy (IDCM) heart failure (HF) patients who will respond to medical therapy (*responders*) and those who will not (*non-responders*) is often similar. A machine learning (ML)-based clinical tool to identify *responders* would prevent unnecessary surgery, while targeting *non-responders* for early intervention. We used regional left ventricular (LV) contractile injury patterns in ML models to identify IDCM HF *non-responders*.

Methods: MRI-based multiparametric strain analysis was performed in 178 test subjects (140 normal subjects and 38 IDCM patients), calculating longitudinal, circumferential, and radial strain over 18 LV sub-regions for inclusion in ML analyses. Patients were identified as *responders* based upon symptomatic and contractile improvement on medical therapy. We tested the predictive accuracy of support vector machines (SVM), logistic regression (LR), random forest (RF), and deep neural networks (DNN).

Results: The DNN model outperformed other models, predicting response to medical therapy with an area under the receiver operating characteristic curve (AUC) of **0.94**. The top features were longitudinal strain in 1) basal: anterior, posterolateral and 2) mid: posterior, anterolateral, and anteroapical sub-regions.

Conclusions: Regional contractile injury patterns predict response to medical therapy in IDCM HF patients, and have potential application in ML-based HF patient care.

*Corresponding Author: Michael K. Pasque, MD, Washington University School of Medicine, Barnes-Jewish Hospital, Division of Cardiothoracic Surgery, Campus Box 8234, 660 S. Euclid Ave., St. Louis, MO 63110, Phone: 314-362-6237, Fax: 314-362-6288, pasquem@wustl.edu.

Keywords

heart failure; regional contractile injury; magnetic resonance imaging; machine learning; deep learning; myocardial strain

Introduction

New-onset idiopathic dilated cardiomyopathy (IDCM) heart failure (HF) patients commonly present in alarming clinical HF with severe congestive symptoms and profound impairment of left ventricular (LV) function, only to subsequently demonstrate an excellent response to medical therapy.¹ Conversely, many patients with an identical clinical presentation do in fact suffer precipitous deterioration.² The inaccuracy of current prognostic metrics often makes the early differentiation between these two groups of patients nearly impossible.

Understandably, the dire consequences of error often result in overreaction, with extensive use of invasive surgery. Under-reaction results in the similarly dire denial of lifesaving surgery to patients who would benefit. A clinical algorithm that combines both clinical and contractile metrics to prospectively identify medical therapy *non-responders* would enable their timely referral for consideration of surgical intervention, while sparing all others.

Machine learning (ML) is optimal for the solution of a wide variety of clinical problems involving predictive modeling.^{3,4,5,6} Unlike most other statistical analyses, ML is capable of informing the investigative process. This is especially applicable in the solution of complex clinical problems, such as the identification of HF medical therapy non-responders. ML informs the investigative process by identifying non-intuitive patterns in variables drawn from widely disparate sources, such as those variables associated with clinical HF. In many cases, the complex distribution patterns of these predictive variables make them undiscoverable by other means. Indeed, the complexity of the pathophysiology associated with clinical HF assures that an equally complex array of variables will be necessary for accurate predictive modeling.

Many *clinical* variables have been correlated to outcomes in HF patients. Unfortunately, clinically applicable metrics associated with poor clinical outcomes in HF, such as ejection fraction, VO₂ max, or cardiac index, do not by themselves demonstrate the predictive accuracy necessary to direct clinical decision making.⁷ Instead of focusing upon single metrics, ML offers the potential to *combine* each of the predictive accuracies of many different variables into a single algorithm to predict response to HF medical therapy on an individual patient basis. Our goal is a ML-based predictive clinical algorithm, and these *clinical* variables, including both directed HF metrics (such as VO₂max) and general electronic health record (EHR) variables, will undoubtedly play a major role in its development.

That being said, the origination of clinical HF is founded in regional LV contractile injury. We have previously demonstrated the predictive potential of high-density regional contractile injury patterns in HF patients.⁸ This initial investigation supplied premise to our hypothesis that regional injury patterns elucidated by a high-density array of *regional* LV

contractile metrics will be foundational in ML models directed toward the recognition of proximity to catastrophic failure in IDCM HF patients.

Since the quality of data input into ML models directly determines their predictive accuracy and clinical efficacy, it is critical that LV contractile function be characterized by accurate, high-density regional contractile metrics. The uniform recognition of cardiac MRI as the gold standard in temporal and spatial resolution of dynamic LV geometry makes it the optimal source for the contractile substrate of predictive HF ML classifiers. Regional LV contractile metrics derived from modalities with inferior spatial resolution, such as echocardiography, are inherently compromised as variables in ML modeling.

In contrast, our MRI-based multiparametric strain analysis derives multiple strain parameter values averaged over an 18-region LV grid from patient-specific cardiac MRI radiofrequency tissue tagging or navigator-gated spiral *Displacement Encoding with Stimulated Echoes* (DENSE)-based strain data.^{8,10} The output from these analyses establishes a comprehensive normalized “digital twin” of each individual patient’s regional LV contractile function. This digital twin uniquely describes, in a highly granular format, the comprehensive contractile functional capability of that patient’s LV in a digital format that is optimal for ML analyses. This digital regional *normalization* of strain is unique to our laboratory and necessary to the high degree of clinical accuracy that we seek, since both normal and pathologic human LV strain is transmurally and regionally heterogeneous. This normalization process also allows the combination of multiple different strain metrics into a multiparametric strain index, which is clinically proven more accurate than its components.¹²

Thus, when strain values are projected over a standardized 15,300-point LV grid, *machine-detection* of regional LV contractile injury can be accomplished by fully-automated, computer-based comparisons of three different patient-specific strain parameter raw values at each of 15,300 LV grid-points (45,900 database comparisons per study) to their respective normal average and standard deviation (SD) by interfacing directly with our laboratory’s *normal human strain database* (n=140 normal volunteers). In ML, this is called *novelty detection*. The machine establishes a *model of normality* and then quantifies and assigns *z-scores* representing the degree of abnormality for each strain component at each LV grid-point in terms of the point-specific standard deviation (SD) from the normal mean.^{9,10,11}

In the setting of these unique clinical investigative capabilities, the goal of this investigation is to test the hypothesis that the contractile injury patterns detected by high-definition regional LV contractile metrics have predictive potential in IDCM patients as ML variables in the identification of non-responders to HF medical therapy.

Materials and Methods

Data Sources

The Washington University School of Medicine Human Studies Institutional Review Board approved this study. Informed and written consent was obtained from normal subjects (n=140) and IDCM HF patients (n=38) who were enrolled in this investigation after recruitment from the medical and surgical HF services at Barnes-Jewish Hospital at

Washington University Medical Center in St. Louis, MO. A consistent team approach utilizing guideline-directed evidence-based HF medical and surgical therapy was carried out in all study patients by the HF services at Washington University Medical Center.

Cardiac MR Image Acquisition and Analysis

Our cardiac MR image acquisition and analysis is described in depth in multiple publications and summarized briefly here.^{12,13} In all cases, imaging was performed in a 1.5 Tesla MR scanner (Sonata, Siemens Medical Systems, Washington D.C.). Short-axis images with grid tagging were acquired in multiple parallel planes extending from the plane of the mitral valve to the apex of the heart. Additionally, four radially oriented long-axis image sets separated by approximately 45 degrees were obtained with line tagging to provide displacement information in the longitudinal direction. In each imaging plane, a spatial modulation of magnetization radiofrequency tissue-tagging preparation was applied, followed by a two-dimensional balanced steady-state free precession cine image acquisition. Typical imaging parameters were repetition time 32.4ms; echo time 1.52ms; field of view 350×350mm; flip angle 20°; tag spacing 8mm; and slice thickness 8mm. DENSE images were obtained using a Navigator-gated 3D spiral DENSE sequence that facilitates volumetric coverage of the entire LV.^{14,15} The spiral k-space trajectory minimizes echo time and increases signal-to-noise ratio compared to existing echo planar techniques.^{14,15} Typical imaging parameters included: repetition time 15msec; echo time 1.10msec; flip angle 15°; FOV 340×340mm; voxel size 2.5×2.5×8.0mm; slice thickness 8mm.

LV Model Construction and Strain Calculation

Manual delineation of the endocardial and epicardial contours was performed at the initial (diastolic) cardiac phase, followed by automatic propagation of these contours to all other cardiac phases. Individual contours from the end-diastolic images were used to construct a spline surface description of the LV at end-diastole. Using the insertion points of the right ventricle into the intraventricular septum as landmarks, an eighteen-region reproducible model of the ventricle was constructed based on AHA guidelines. For the tagged data, tag lines were tracked interactively on each image using an active contour approach.¹⁶ Systolic displacements were computed at the intersection points of the three tag surfaces. In the case of DENSE images, displacements were computed at each voxel within the segmented myocardium using a phase unwrapping algorithm.¹⁷

Regional values of circumferential, longitudinal, and radial strain were computed in each patient using the radial point interpolation method (RPIM) as described previously.¹⁸ These regional strain measurements served as input variables for the ML analysis. As previously mentioned, a fully-automated comparison of three different patient-specific strain parameter raw values at each of 15,300 LV grid-points (45,900 database comparisons per study) is made to their respective normal average and standard deviation (SD) by interfacing directly with our laboratory's *normal human strain database* (n=140 normal volunteers). In this *novelty detection* process, the machine establishes a *model of normality* from the data obtained from the 140 normal volunteers. It then compares each patient's raw strain data in a point-by-point, component-by-component fashion to its respective normal values with calculation of *z-scores* for each component at each point. The *z-scores* represent the degree

of abnormality for each strain component at each LV grid-point in terms of the component-specific/point-specific standard deviation (SD) from the normal mean.^{9,10,11} This normalization process enables consistent, reproducible, and objective machine detection of patient-specific regional contractile injury patterns, as well as optimizing clinical utility by expression in an intuitive fashion that allows temporal and intergroup comparisons.

Statistical Analysis

Select demographic and clinical variables were compared between IDCM HF patient *responders* (n=29) and *non-responders* (n=9). Continuous variables were expressed as mean $\pm$ standard deviation, as shown in Table 1. Student's t-test was used to compare means of normally distributed continuous variables, while Mann-Whitney *U* test was used for skewed distributions. Categorical variables were compared using Fisher's Exact test. P-value < 0.05 was considered statistically significant.

Machine Learning Analysis

To enable supervised ML analyses with binary outcome classification, each patient was classified as either a "responder" or "non-responder" based upon the presence or absence of *sustained* symptomatic [> 1 New York Heart Association (NYHA) class] and contractile ($>10\%$ increase echocardiography-based ejection fraction) improvement over the follow-up period. This classification was carried out by two HF specialists who were blinded to the results of the MRI-based multiparametric strain contractile analysis. Echocardiography was employed in the *clinical* classification of patients during the follow-up period because of its availability in all of the study HF patients.

We first investigated the feature importance of 54 strain variables by a tree-based classifier on the complete dataset. Then, we tested ML and deep learning algorithms, including support vector machines (SVM), logistic regression (LR), random forest (RF), and deep neural networks (DNN) and repeated each 54 times by using the first top *n* variables ($n = 54, 53, 52, \dots, 1$). For each iteration, we randomly divided the dataset into 4 folds, and trained the models using 3-folds as the training data, and evaluated the models using one-fold as the testing data. We repeated this 4-fold cross-validation repeated 10 times to obtain the mean and standard deviation of model performance metrics. The Synthetic Minority Oversampling Technique (SMOTE) was employed to produce balanced training sets.

All ML analyses were carried out using Python (www.python.org) programming language. Data was prepared and managed using scikit-learn (<https://scikit-learn.org>), NumPy (<https://numpy.org>), pandas (<https://pandas.pydata.org>), TensorFlow (<https://www.tensorflow.org>), and Keras (<https://keras.io>).

Machine Learning and Deep Learning Analysis

The DNN model was composed of an input layer of features followed by 5 hidden layers (each layer having 256, 256, 128, 64, and 32 neurons, respectively) and a scalar output layer. We used binary cross-entropy as the loss function of the output layer and employed the ReLu function⁹ as the nodal activation function for each hidden layer. We used the Adam optimizer¹⁰ and a mini-batch size of 64. This final model was obtained by the iterative

testing of multiple variations using different numbers of nodal dimensions, as well as multiple different nodal activation functions, loss functions, and optimizers. The final model resulted in the best DNN predictive accuracy.

The RF model was configured as follows: the number of trees in the random forest was set 100; the number of maximum features can be used in each tree was set as the square root of the total number of features; the minimum number of samples at a leaf node of a tree was set as 1. The logistic regression model was configured as follows: the L2 norm was used in the penalization, i.e., the variance of predicted value and real value of training data; the stopping criteria was set as 1.0×10^{-4} ; the inverse of regularization strength, which reduces the potential overfitting, was set as 1.0. The SVM model was configured as follows: the regularization parameter C, which regulates the balance between correct classification of training samples and maximization of the classification margin, was set as 1.0; the radial basis function (RBF) kernel, to calculate the distance of two samples, was used; the kernel coefficient gamma, which affects the inverse radius of the support vectors, was set as $1/(\text{number of features} \times \text{variance of training data})$. Feature importance analysis indicates feature value in the construction of boosted decision trees within the model. The frequency of feature use determines its relative importance.

Results

Baseline Characteristics

There were no differences in gender ($P=1.00$), BMI ($P=0.794$), and rates of diabetes ($P=0.673$) or New York Heart Association Class III/IV HF ($P=0.147$) between the IDCM HF patients who responded to medical therapy and those who did not (Table 1). In addition, there were no differences in serum creatinine ($P=0.905$), serum BUN ($P=0.306$), serum sodium ($P=0.546$), and serum albumin ($P=0.459$) between the two cohorts. Though not statistically significant, the IDCM *non-responders* were older (51.4 ± 12.2 years vs 44.0 ± 9.76 years, $P=0.106$) and had a lower baseline left ventricular ejection fraction ($21.9 \pm 7.9\%$ vs $28.4 \pm 13.2\%$, $P=0.168$), compared to the patients who responded to therapy. As expected, the post therapy ejection fraction was significantly better in the responder group compared to the non-responders ($48.1 \pm 11.1\%$ versus $25.5 \pm 15.0\%$; respectively; p value = 0.002). It is readily apparent that not all *responders* necessarily returned to a normal ($>50\%$) ejection fraction. The *non-responders* also had higher indexed end-diastolic diameter (3.6 ± 0.5 vs 3.1 ± 0.6 , $P=0.016$) and indexed end-systolic diameter (3.2 ± 0.6 vs 2.6 ± 0.7 , $P=0.014$).

Feature Importance Analysis

The top 5 features were longitudinal strain components in 1) basal: anterior, posterolateral and 2) mid: posterior, anterolateral, and anteroseptal sub-regions (Figure 1). The feature importance analysis was utilized to optimize the solution of the ML models, as described below.

Receiver Operating Characteristic Curve Analysis

Table 2 demonstrates model performances for DNN, SVM, LR, and RF by AUC values. The first column demonstrates the models based upon the number of the most important features

utilized, in order of importance. The optimal model was a DNN model utilizing the top 37 features, as identified by tree-based classifier. It predicted response to medical therapy with AUC of **0.94**.

The ROC curves of the 4-folds of the DNN model utilizing the optimal 37 features identified by tree-based classifier are demonstrated in Figure 2. The SVM, logistic regression, and random forest models demonstrated a maximum AUC of 0.79, 0.79, and 0.86 by using the optimal 11, 9, and 5 features, respectively. The SVM model achieved an AUC of 0.67 by using all 54 features, and an AUC of 0.79 by using only top 8 to 11 features. Similarly, LR had an AUC of 0.70 by using all 54 features and 0.79 with the top 9 features; RF had an AUC of 0.7 with 54 features and 0.86 by using the top 5 features. *The results indicated the usefulness of feature selection analysis for the non-DNN models. For these non-DNN models, excessive features can result in overfitting on the training data, and consequently have relatively limited generalization performance on the testing data. The DNN model is not sensitive to the feature selection, and achieved almost the best performance with using all the features.*

Other metrics are displayed in Table 1 of Supplementary Material, including the overall accuracy, precision, recall, and F1-score for the DNN model. Definitions of the metrics are as follows: Accuracy = (true positive + true negative) / (true positive + true negative + false positive + false negative); Precision = true positive / (true positive + false positive); Recall = true positive / (true positive + false negative); F1-score = 2 * (Precision * Recall) / (Precision + Recall).

Discussion

Since HF in IDCM patients occurs because myocyte contractile dysfunction impairs the ability of the heart to pump blood, metrics of regional contractile function offer potential to provide clinically applicable prognostic variables. The qualitative global contractile indices currently being clinically utilized, including such parameters as ejection fraction, cardiac index, ventricular filling pressures, and fractional shortening indices, are insensitive on an individual-patient basis to the subtle group differences that would allow *individual patient-specific* risk stratification in this heterogeneous HF population. In contrast to these qualitative indices, this preliminary investigation suggests *regional distribution patterns* of MRI-based *machine-detected* regional contractile injury have potential to accurately predict clinical outcomes in IDCM HF patients. Preliminary DNN ML analyses using *only* high-density regional LV strain metrics (LV contractile *digital twin*) from 38 IDC HF patients demonstrated an accuracy of 85% in discriminating medical therapy responders from non-responders with an AUC of **0.94**.

Our results establish the use of patient-specific, LV contractile *digital twins* in the predictive management of IDCM HF patients. Each HF patient digital twin representation, encompassed by the patient-specific values of their average regional strain indices (n=54) is ideally suited to immediate interrogation by ML models. It can be likened to the implementation and ML interrogation of the digital twin of modern-day jet engines, each having several thousand continuously monitored sensors.¹⁹ The normal range of each of our

strain metrics is known (our *normal human strain database*), just like the normal value ranges of each of the jet engine sensors. The digital twin of each jet engine is updated on a daily basis by downloading terabytes of data into maintenance servers where this information is interrogated by advanced ML models trained to recognize sensor data patterns that indicate proximity to known engine failure modes. In every airline across the globe, ML-based detection of jet engine sensor output patterns indicative of proximity to a failure mode automatically leads to immediate removal of the engine from service, and triggers a ML-based diagnostic and repair algorithm. Just as advanced ML models can predict proximity to impending jet engine failure by identifying predictive sensor output patterns, so also, we hypothesize, can the ML model developed to interrogate the digital twin of the failing human heart. So also, we hypothesize, can ML-based identification of LV strain metric patterns indicative of proximity to catastrophic LV decompensation assist in triggering HF surgical therapeutic algorithms. This strategy of ML monitoring for regional contractile injury patterns that are predictive of poor response to continued medical therapy would augment the current clinical paradigm. Non-responders who have the most to gain from consideration of high-risk HF surgical intervention with mechanical ventricular assistance or cardiac replacement therapy can be accurately identified before life-ending cardiac decompensation.

Many other diagnostic and therapeutic clinical capabilities accompany the availability of a patient-specific, contractile digital twin based in the *digitally defined* 3D space encompassed by our high-density LV grid. Perhaps the most useful is its ability to concisely summarize patient *subpopulation* regional LV contractile injury patterns into a single intuitive visual compilation of “whole group” contractile characteristics. Because of the standardized and reproducible uniformity of each patient’s LV grid, this is accomplished by averaging all of the respective point-specific values from each of the group’s constituent patient *digital twins*. The resulting average multiparametric strain z-scores are then plotted over LV geometry. Figure 3 represents such a summary compilation obtained by averaging point-specific multiparametric strain data from both medical therapy 1) *non-responders* (left) and 2) *responders* (right) in a subset of our patients. Sub-group average “normalized” contractile function is indicated by standard color contour mapping of group average z-scores with red, yellow, and blue areas representing ≥ 2 , < 2 -1 , and < 1 standard deviations from the normal mean contractile function. These group composite LV maps demonstrate the clear differences between those patients who would have a good response to medical therapy (right) and those who would not (left). The intense concentration of contractile injury centered in the *non-responders*’ septum is readily apparent. Summary LV plots can enhance data analysis and clinical applicability by graphically representing subgroup regional contractile injury patterns.

Another unique capability enabled by patient-specific LV contractile *digital twins* is the generation of 3D strain “subtraction” maps that quantify and localize on a regional basis any *changes* in normalized strain that occur 1) across any clinical interval or intervention (such as changes in medical therapy); or 2) between groups. An example of a “subtraction” image drawn from our preliminary test groups is shown in Figure 4. In a subset of our study group, we subtracted, on a point-by-point basis, the average summated digital twin of the *responder* group from that of the *non-responder* group and used standard 3D color contour mapping

algorithms to plot the results. Figure 4, therefore, comprises *4-quadrant views* of the resulting composite 3D LV subtraction map. Of particular interest, red areas represent >1.0 SD worse average contractile injury in the *non-responders*, while blue regions demonstrated no differences in injury between the groups. This unique 3D LV plot therefore demonstrates, in an intuitive graphical format, the regional contractile *differences*—quantified in standard deviations from the normal mean—between those patients who will ultimately demonstrate a positive response to medical therapy, and those who will not.

Perhaps even more important to the HF clinician, this subtraction mapping can also be performed in the individual HF patient over follow-up intervals. For instance, baseline *digital twin* data can be subtracted from a repeat *digital twin* obtained at follow up after changes in their HF medical regimen, uniquely demonstrating the precise changes in normalized regional contractile function resulting from the change in medical therapy.

Finally, our analysis showed the top five features used to optimize the solution of the ML models were longitudinal strain components in 1) basal: anterior, posterolateral, and 2) mid: posterior, anterolateral, and antero-septal sub-regions. This is consistent with our previous studies that demonstrated the septum to be a “sentinel” injury zone characterized by early and severe contractile injury in dilated cardiomyopathy patients, with relative sparing of the lateral walls⁸. The ML feature importance results reported in the current study are, therefore, not surprising since our previous study suggested that a lack of injury in the septum portended a near-uniform good outcome, while severe injury in the lateral wall was almost always associated with a bad outcome. The predominance of longitudinal strain components in the classification of outcomes is further evidence for its importance in normal and pathological contractile function and emphasizes the importance of a full multiparametric strain analysis in the development of regional contractile-based machine learning predictive modeling in dilated cardiomyopathy heart failure.

Limitations

Our study has several limitations. First, we utilize cardiac MRI rather than echocardiography, which is the current clinical standard. Echocardiography holds this position secondary to its ease of application and ready availability in most acute care clinical settings. That being said, our MRI scan time (<20 min) is now less than an average echocardiogram and delays in echocardiogram interpretation are obviated by our fully automated contractile injury detection algorithm (<20 sec). Although MRI costs still significantly exceed those of echocardiography, the differences in cost are shrinking—while MRI continues to offer far superior spatial resolution.²⁰ Since ML predictive accuracy, given equally effective algorithms, is entirely data quality-driven, cardiac MRI is the optimal choice. Further limitations are found in the relatively small number of non-responder IDCM HF patients ($n=9$), thereby increasing the likelihood of a type II statistical error when comparing the two groups. Despite this small number of non-responders, the deep learning model demonstrated considerable classification power in a 4-fold cross-validation repeated 10 times, generating an AUC of 0.94 using only 37 features. This, in addition to the fact that all patients included in this analysis were recruited from a single institution, may impair the extrapolation of our ML results to other centers. Finally, we investigated the feature

importance of 54 strain variables by a tree-based classifier in the *complete* dataset which might result in overestimation of the final model performance. We will investigate more approaches to deal with this in our future work.

In conclusion, the clinical impact on HF patient populations of a hybrid (contractile/clinical variables) ML model that accurately prospectively differentiates *responders* from *non-responders* to HF medical therapy would be profound. New-onset HF patients who would be predicted to have severe progression of HF on medical therapy alone and, therefore, a high risk of HF adverse events (catastrophic progression of HF), would be directed toward surgical intervention after initiation of intense medical therapy, but before end-organ damage or sudden life-ending cardiac events occur. In contrast, those who can be reliably predicted to have a good response to medical therapy and a low incidence of HF adverse events would avoid unnecessary surgical intervention. The lack of an accurate risk stratification strategy is the greatest limitation to the *accurate* application of surgical intervention in HF patients. A clinical tool with the accuracy to direct patient-specific therapeutic intervention at initial HF presentation is an achievable goal and high-density metrics of regional LV contractile injury have considerable potential to contribute to these future ML clinical algorithms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors have no conflicts of interest to disclose. Relevant funding sources included National Institutes of Health 1R01HL112804 and 1R56HL136619.

Abbreviations and Acronyms

AUC	area under the curve
DENSE	<i>Displacement Encoding with Stimulated Echoes</i>
DNN	deep neural networks
EHR	electronic health record
HF	heart failure
IDCM	idiopathic dilated cardiomyopathy
LR	logistic regression
LV	left ventricle
ML	machine learning
NYHA	New York Heart Association
RF	random forest
ROC	receiver operating characteristic curve

SVM support vector machines

References

1. Merlo M, Cannata A, Gobbo M, et al. Evolving concepts in dilated cardiomyopathy. *Eur J Heart Fail*. 2018;20(2):228–39. [PubMed: 29271570]
2. Jefferies J, Towbin J. Dilated cardiomyopathy. *The Lancet*. 2010;375(9716):752–62.
3. Churpek M, Yuen T, Winslow C, et al. Multicenter Comparison of Machine Learning Methods and Conventional Regression for Predicting Clinical Deterioration on the Wards. *Crit Care Med*. 2016;44:368–74. [PubMed: 26771782]
4. Taylor R, Pare J, Venkatesh A, et al. Prediction of In-hospital Mortality in Emergency Department Patients with Sepsis: A Local Big Data-Drive, Machine Learning Approach. *Acad Emerg Med Off J Soc Acad Emerg Med*. 2016;23:269–278.
5. Liu N, Holcomb J, Wade C, et al. Utility of vital signs, heart rate variability and complexity, and machine learning for identifying the need for lifesaving interventions in trauma patients. *Shock (Augusta Ga)*. 2014;42:108–114.
6. Wojnarski C, Roselli E, Idrees J, et al. Machine-learning phenotypic classification of bicuspid aortopathy. *J Thorac Cardiovas Surg*. 2017;155(2):461–69.
7. Nohria A, Tsang S, Fang J, et al. Clinical assessment identifies hemodynamic profiles that predict outcomes in patients admitted with heart failure. *Journal of the American College of Cardiology*. 2003;41(10):1797–1804. [PubMed: 12767667]
8. Henn M, Lawrance C, Kar J, et al. Dilated Cardiomyopathy: Normalized Multiparametric Myocardial Strain Predicts Contractile Recovery. *The Annals of Thoracic Surgery*. 2015;100(4):1284–91. [PubMed: 26228597]
9. Joseph S, Moazami N, Cupps BP, Howells A, Craddock H, Ewald G, Rogers J, Pasque MK. Magnetic resonance imaging-based multiparametric systolic strain analysis and regional contractile heterogeneity in patients with dilated cardiomyopathy. *J Heart Lung Transplant*. 2009;28(4):388–94. doi: 10.1016/j.healun.2008.12.018. [PubMed: 19332267]
10. Lancaster TS, Kar J, Cupps BP, Henn MC, Kulshrestha K, Koerner DJ, Pasque MK. Topographic mapping of left ventricular regional contractile injury in ischemic mitral regurgitation. *The Journal of thoracic and cardiovascular surgery*. 2017;154(1):149–58 e1. doi: 10.1016/j.jtcvs.2016.11.055. [PubMed: 28109612]
11. Naylor CD. On the Prospects for a (Deep) Learning Health Care System. *JAMA*. 2018;320(11):1099–100. doi: 10.1001/jama.2018.11103. [PubMed: 30178068]
12. Cupps BP, Bree DR, Wollmuth JR, Howells AC, Voeller RK, Rogers JG, Pasque MK. Myocardial viability mapping by magnetic resonance-based multiparametric systolic strain analysis. *The Annals of thoracic surgery*. 2008;86(5):1546–53. [PubMed: 19049746]
13. Cupps BP, Taggar AK, Reynolds LM, Lawton JS, Pasque MK. Regional myocardial contractile function: multiparametric strain mapping. *Interact Cardiovasc Thorac Surg*. 2010;10(6):953–7. [PubMed: 20233804]
14. Spottiswoode BS, Zhong X, Hess AT, Kramer CM, Meintjes EM, Mayosi BM, Epstein FH. Tracking myocardial motion from cine DENSE images using spatiotemporal phase unwrapping and temporal fitting. *IEEE Trans Med Imaging*. 2007;26(1):15–30. [PubMed: 17243581]
15. Zhong X, Spottiswoode BS, Meyer CH, Kramer CM, Epstein FH. Imaging three-dimensional myocardial mechanics using navigator-gated volumetric spiral cine DENSE MRI. *Magn Reson Med*. 2010;64(4):1089–97. [PubMed: 20574967]
16. Moulton M, Cresswell L, Downing S, et al. Spline surface interpolation for calculating 3D ventricular strains from MRI tissue tagging. *Am J Physiol*. 1996;39:H281–97.
17. Kar J, Knutsen AK, Cupps BP, Pasque MK. A validation of two-dimensional in vivo regional strain computed from displacement encoding with stimulated echoes (DENSE), in reference to tagged magnetic resonance imaging and studies in repeatability. *Ann Biomed Eng*. 2014;42(3):541–54. [PubMed: 24150239]
18. Kar J, Knutsen AK, Cupps BP, Zhong X, Pasque MK. Three-dimensional regional strain computation method with displacement encoding with stimulated echoes (DENSE) in non-

ischemic, non-valvular dilated cardiomyopathy patients and healthy subjects validated by tagged MRI. *J Magn Reson Imaging*. 2015;41(2):386–96. [PubMed: 24753028]

19. Tarassenko L, Topol EJ. Monitoring Jet engines and the Health of People. *JAMA*. 2018;320(22):2309–10. [PubMed: 30453330]
20. Gardner B, Bingham S, Allen M, et al. Cardiac magnetic resonance versus transthoracic echocardiography for the assessment of cardiac volumes and regional function after myocardial infarction: an intrasubject comparison using simultaneous intrasubject recordings. *Cardiovascular Ultrasound*. 2009;7(38):1–7. [PubMed: 19128462]

Figure 2.

The Receiver Operating Characteristic curves of the 4-folds cross-validation repeated 10 times in the DNN model utilizing the optimal 37 features. DNN = deep neural networks.

Figure 3.

Summary compilation obtained by averaging point-specific multiparametric strain data from both medical therapy *non-responders* (left) and *responders* (right). Standard color contour algorithms display group average 3D regional contractile injury by standard deviation (comparison to normal human strain database): red, yellow, and blue areas representing ≥ 2 , < 2 ≥ 1 , and < 1 standard deviations from the normal mean, respectively. Presentation in standard deviations from the normal mean allows a familiar and clinically relevant calibration of these changes. P = posterior; PS = posteroseptal; AS = anteroseptal; A = anterior; AL = anterolateral; PL = posterolateral.

Figure 4.

4-quadrant views of a composite 3D left ventricular “subtraction” map. The group average “normalized” *digital twin* data (expressed in standard deviations from the normal mean; z-scores) from the *non-responders* was subtracted on a point-by-point basis from the *responders*, and the results were plotted using standard color contour mapping. On this LV map, red areas represent >1.0 SD worse average contractile injury in the *non-responders*, while blue regions demonstrated no differences in injury between the groups. Once again, the standard deviation presentation allows a clinically relevant calibration of the differences between the two subgroups. P = posterior; PS = posteroseptal; AS = anteroseptal; A = anterior; AL = anterolateral; PL = posterolateral.

Table 1.

Demographic and Clinical Variables

Variable	IDCM HF Patient Responders (n = 29)	IDCM HF Patient Non-Responders (n = 9)	P value
Age (years), mean $\pm$ std	44.0 $\pm$ 9.76	51.4 $\pm$ 12.2	0.106
Male gender, n (%)	17 (59%)	6 (67%)	1.000
BMI (kg/m ²), mean $\pm$ std	28.6 $\pm$ 6.4	27.9 $\pm$ 6.2	0.794
Diabetes, n (%)	7 (24%)	3 (33%)	0.673
NYHA Class III/IV, n (%)	6 (27%)	4 (67%)	0.147
Serum creatinine, mean $\pm$ std	1.0 $\pm$ 0.2	1.1 $\pm$ 0.3	0.905
Serum BUN, mean $\pm$ std	19.4 $\pm$ 6.8	16.8 $\pm$ 5.3	0.306
Serum sodium, mean $\pm$ std	139.0 $\pm$ 2.7	138.3 $\pm$ 2.8	0.546
Serum albumin, mean $\pm$ std	3.9 $\pm$ 0.5	3.7 $\pm$ 0.7	0.459
Baseline LVEF, mean $\pm$ std	28.4 $\pm$ 13.2	21.9 $\pm$ 7.9	0.168
Latest LVEF, mean $\pm$ std	48.1 $\pm$ 11.1	25.5 $\pm$ 15.0	0.002
EDD/BSA, mean $\pm$ std	3.1 $\pm$ 0.6	3.6 $\pm$ 0.5	0.016
ESD/BSA, mean $\pm$ std	2.6 $\pm$ 0.7	3.2 $\pm$ 0.6	0.014
PA systolic pressure, mean $\pm$ std	35.9 $\pm$ 10.7	37.3 $\pm$ 12.8	0.773

BMI = body mass index; BSA = body surface area (m²); BUN = blood urea nitrogen; EDD = end diastolic diameter; ESD = end systolic diameter; LVEF = left ventricular ejection fraction; NYHA = New York Heart Association

Table 2.

DNN, SVM, LR, RF Average Performance by 4-fold Cross Validation Repeated 10 times.

Total # of features used	DNN AUC	SVM AUC	LR AUC	RF AUC
54	0.93 (0.05)	0.67 (0.02)	0.7 (0.04)	0.7 (0.09)
53	0.92 (0.04)	0.66 (0.02)	0.68 (0.04)	0.66 (0.04)
52	0.91 (0.04)	0.66 (0.02)	0.7 (0.03)	0.77 (0.04)
51	0.91 (0.06)	0.66 (0.01)	0.7 (0.03)	0.7 (0.06)
50	0.91 (0.05)	0.66 (0.02)	0.7 (0.03)	0.69 (0.04)
49	0.88 (0.08)	0.67 (0.01)	0.71 (0.04)	0.7 (0.08)
48	0.92 (0.03)	0.66 (0.02)	0.7 (0.04)	0.73 (0.05)
47	0.93 (0.03)	0.66 (0.02)	0.71 (0.03)	0.7 (0.06)
46	0.92 (0.05)	0.66 (0.01)	0.71 (0.04)	0.7 (0.05)
45	0.93 (0.04)	0.66 (0.01)	0.69 (0.04)	0.7 (0.06)
44	0.92 (0.06)	0.66 (0.01)	0.68 (0.03)	0.68 (0.07)
43	0.93 (0.05)	0.66 (0.01)	0.7 (0.03)	0.69 (0.1)
42	0.92 (0.08)	0.65 (0.02)	0.71 (0.04)	0.73 (0.04)
41	0.92 (0.05)	0.67 (0.02)	0.72 (0.02)	0.72 (0.05)
40	0.92 (0.05)	0.66 (0.02)	0.72 (0.03)	0.72 (0.08)
39	0.93 (0.05)	0.67 (0.02)	0.7 (0.04)	0.71 (0.06)
38	0.92 (0.05)	0.66 (0.02)	0.71 (0.03)	0.73 (0.05)
37	0.94 (0.03)	0.66 (0.01)	0.72 (0.03)	0.7 (0.06)
36	0.93 (0.04)	0.66 (0.02)	0.7 (0.03)	0.72 (0.05)
35	0.92 (0.03)	0.65 (0.01)	0.71 (0.02)	0.74 (0.08)
34	0.91 (0.03)	0.65 (0.03)	0.71 (0.02)	0.75 (0.05)
33	0.92 (0.06)	0.65 (0.02)	0.7 (0.03)	0.72 (0.09)
32	0.91 (0.04)	0.65 (0.01)	0.71 (0.03)	0.68 (0.08)
31	0.93 (0.03)	0.65 (0.02)	0.71 (0.03)	0.71 (0.08)
30	0.89 (0.07)	0.65 (0.03)	0.7 (0.04)	0.72 (0.07)
29	0.91 (0.04)	0.65 (0.01)	0.7 (0.04)	0.7 (0.11)
28	0.9 (0.04)	0.66 (0.02)	0.71 (0.03)	0.75 (0.05)
27	0.9 (0.03)	0.65 (0.02)	0.71 (0.03)	0.74 (0.05)
26	0.93 (0.04)	0.7 (0.06)	0.72 (0.02)	0.76 (0.05)
25	0.93 (0.03)	0.73 (0.04)	0.72 (0.02)	0.73 (0.05)
24	0.92 (0.06)	0.74 (0.03)	0.7 (0.04)	0.73 (0.06)
23	0.93 (0.04)	0.75 (0.02)	0.71 (0.02)	0.73 (0.07)
22	0.93 (0.05)	0.75 (0.02)	0.72 (0.03)	0.74 (0.04)
21	0.93 (0.02)	0.75 (0.03)	0.71 (0.02)	0.76 (0.08)
20	0.93 (0.04)	0.74 (0.03)	0.72 (0.02)	0.76 (0.05)
19	0.91 (0.03)	0.76 (0.01)	0.73 (0.05)	0.77 (0.08)
18	0.93 (0.04)	0.76 (0.01)	0.71 (0.04)	0.76 (0.07)

Total # of features used	DNN AUC	SVM AUC	LR AUC	RF AUC
17	0.9 (0.06)	0.75 (0.02)	0.72 (0.03)	0.79 (0.06)
16	0.9 (0.03)	0.75 (0.01)	0.72 (0.01)	0.77 (0.06)
15	0.91 (0.04)	0.75 (0.02)	0.74 (0.02)	0.78 (0.05)
14	0.91 (0.02)	0.76 (0.03)	0.73 (0.02)	0.76 (0.05)
13	0.92 (0.02)	0.78 (0.03)	0.74 (0.02)	0.77 (0.05)
12	0.91 (0.02)	0.77 (0.04)	0.73 (0.02)	0.78 (0.07)
11	0.9 (0.04)	0.79 (0.02)	0.74 (0.03)	0.82 (0.06)
10	0.89 (0.03)	0.79 (0.02)	0.76 (0.04)	0.83 (0.06)
9	0.91 (0.02)	0.79 (0.01)	0.79 (0.03)	0.83 (0.05)
8	0.9 (0.04)	0.79 (0.03)	0.77 (0.02)	0.79 (0.04)
7	0.92 (0.02)	0.76 (0.03)	0.77 (0.03)	0.83 (0.05)
6	0.92 (0.03)	0.76 (0.01)	0.76 (0.02)	0.81 (0.06)
5	0.88 (0.04)	0.77 (0.01)	0.7 (0.04)	0.86 (0.04)
4	0.88 (0.04)	0.75 (0.01)	0.73 (0.03)	0.82 (0.05)
3	0.87 (0.03)	0.75 (0.02)	0.77 (0.04)	0.82 (0.05)
2	0.84 (0.03)	0.72 (0.01)	0.7 (0.02)	0.68 (0.06)
1	0.82 (0.05)	0.76 (0.01)	0.75 (0.02)	0.63 (0.05)

DNN = deep neural networks; SVM = support vector machines; LR = logistic regression; RF = random forests; AUC = area under receiver operating curve.