

RESEARCH

Open Access



Neural-network classification of cardiac disease from ^{31}P cardiovascular magnetic resonance spectroscopy measures of creatine kinase energy metabolism

Meiyappan Solaiyappan¹, Robert G. Weiss² and Paul A. Bottomley^{1*} 

Abstract

Background: The heart's energy demand per gram of tissue is the body's highest and creatine kinase (CK) metabolism, its primary energy reserve, is compromised in common heart diseases. Here, neural-network analysis is used to test whether noninvasive phosphorus (^{31}P) cardiovascular magnetic resonance spectroscopy (CMRS) measurements of cardiac adenosine triphosphate (ATP) energy, phosphocreatine (PCr), the first-order CK reaction rate k_f , and the rate of ATP synthesis through CK (CK flux), can predict specific human heart disease and clinical severity.

Methods: The data comprised the extant 178 complete sets of PCr and ATP concentrations, k_f , and CK flux data from human CMRS studies performed on clinical 1.5 and 3 Tesla scanners. Healthy subjects and patients with nonischemic cardiomyopathy, dilated (DCM) or hypertrophic disease, New York Heart Association (NYHA) class I-IV heart failure (HF), or with anterior myocardial infarction are included. Three-layer neural-networks were created to classify disease and to differentiate DCM, hypertrophy and clinical NYHA class in HF patients using leave-one-out training. Network performance was assessed using 'confusion matrices' and 'area-under-the-curve' (AUC) analyses of 'receiver operating curves'. Possible methodological bias and network imbalance were tested by segregating 1.5 and 3 Tesla data, and by data augmentation by random interpolation of nearest neighbors, respectively.

Results: The network differentiated healthy, HF and non-HF cardiac disease with an overall accuracy of 84% and AUC > 90% for each category using the four CK metabolic parameters, alone. HF patients with DCM, hypertrophy, and different NYHA severity were differentiated with ~ 80% overall accuracy independent of CMRS methodology.

Conclusions: While sample-size was limited in some sub-classes, a neural network classifier applied to noninvasive cardiac ^{31}P CMRS data, could serve as a metabolic biomarker for common disease types and HF severity with clinically-relevant accuracy. Moreover, the network's ability to individually classify disease and HF severity using CK metabolism alone, implies an intimate relationship between CK metabolism and disease, with subtle underlying phenotypic differences that enable their differentiation.

Trial registration: ClinicalTrials.gov Identifier: NCT00181259.

Keywords: Translational studies, Cardiac metabolism, Biomarker, Neural network, Heart failure, Phosphorus spectroscopy

* Correspondence: bottoml@mri.jhu.edu

¹Division of MR Research, Department of Radiology, Johns Hopkins School of Medicine, Park Bldg. 310, 600 N Wolfe St, Baltimore, MD 21287, USA
Full list of author information is available at the end of the article



Background

Since the mid-1980s numerous noninvasive localized phosphorus (^{31}P) cardiovascular magnetic resonance spectroscopy (CMRS) studies have been performed on patients with heart disease to measure key energy metabolites—phosphocreatine (PCr) and adenosine triphosphate (ATP) [1]. ATP is the source of the cellular energy that fuels all cellular processes including myocardial contraction. PCr is the heart's primary energy reserve, producing ATP from the phosphorylation of adenosine diphosphate (ADP) via the creatine kinase (CK) reaction: $\text{PCr} + \text{ADP} \rightleftharpoons \text{Cr} + \text{ATP}$, forming creatine (Cr) as a byproduct. The majority of human cardiac ^{31}P CMRS studies use the PCr/ATP ratio as an index of this cardiac energy reserve [1]. Significant reductions in myocardial PCr/ATP are reported for a wide range of cardiac diseases including acute ischemia [2], dilated cardiomyopathy (DCM) [3–6], hypertrophic cardiomyopathy (HCM) [7–11], heart failure (HF) [3–11], valve disease [7, 8], obesity [12], diabetes [13] and in patients with heart transplants [14]. Although there is a trend for greater reductions in worse illness [4, 5, 8–10], the scatter and overlap amongst PCr/ATP values has nevertheless precluded its use as a metabolic biomarker for any specific condition or for grading disease severity, at least on an individual basis [1].

In the 1990s, methodologies that combine internal or external concentration references with magnetic resonance imaging (MRI) volumetry [10, 15–18], have enabled the measurement of absolute concentrations of human cardiac creatine kinase (CK) metabolites, revealing deficits in patients with cardiomyopathy, HF and myocardial infarction (MI) [6, 10, 11, 17]. In addition, the development of fast ^{31}P saturation-transfer techniques over the past decade have provided measurements of the pseudo first-order kinetic reaction rate, k_f (in s^{-1}) for the CK reaction [19, 20]. Reduced myocardial k_f was observed in patients with left ventricular hypertrophy (LVH) and HF, but not in those without HF [11]. This suggested that reductions in CK k_f might be specific to HF. By combining k_f and PCr concentration ([PCr]) measurements in the same cardiac ^{31}P CMRS exam, the forward CK flux of ATP supplied to the heart can be obtained [6]. This is given by the product, k_f [PCr].

In nonischemic cardiomyopathy (NICM), CK flux declines with progression to HF [1, 21]. Indeed, the reduction in CK flux independently predicted cardiac events (HF hospitalization, transplantation, left-ventricular (LV) assist device requirement, cardiac death) in HF patients that occurred several years post-CMRS [22]. More recently it was shown that CK flux correlates with cardiac mechanical work [23], providing a mechanistic link between declining CK flux, mechanical work and cardiac events. While these findings support

the 'energy starvation hypothesis' for HF [24, 25] and suggest that such measurements might index disease severity, the scatter among individual CK metabolite concentrations, k_f , and flux data, again preclude the discrimination of disease type or severity on an individual basis, using any one ^{31}P CMRS metric alone.

Today, trained machine learning techniques employing multi-layer neural networks are emerging as powerful tools for identifying and classifying a vast array of multi-parametric input data in applications ranging from bioinformatics to sociology to driving cars. Once designed, a neural network can be retuned and retrained as additional data become available, without exhaustively searching for the best classifier. Here, we apply neural-network analysis to test whether a combination of cardiac [PCr], [ATP], k_f and the CK flux supplying ATP to the heart, as measured noninvasively with ^{31}P CMRS, can distinguish several common myocardial diseases and their clinical severity. The study includes—to the best of our knowledge—the entire known data pool of subjects for which complete sets of [PCr], [ATP], k_f and CK flux measurements exist [6, 11, 21–23, 26, 27]. The work tests the hypothesis that cardiac CK metabolism is altered in patterns that may be specific to different forms of heart disease and/or to the clinical severity of disease, specifically in HF. If true, cardiac CK metabolic data acquired with ^{31}P CMRS could be used with a neural network classifier as a metabolic biomarker to gauge disease type and severity. Moreover, the existence of such relationships would link certain common clinical cardiac conditions to compromised CK energy supply in individual patients.

Methods

Subjects

Complete sets of cardiac [PCr], [ATP], k_f and CK flux data were obtained from 178 noninvasive cardiac ^{31}P CMRS saturation transfer studies performed between 1999 and 2011 acquired with the approval of the Johns Hopkins University Institutional Review Board [6, 11, 21–23, 26, 27]. All participants provided written informed consent. Forty-five of the studies were from subjects 19–65 years-old (mean 40 ± 12 years) who were considered healthy with no history of heart disease, diabetes, or hypertension (controls). There were 109 patients, 20–72 years-old (mean 47 ± 12 years), with NICM including some with LVH who had no evidence of critical coronary disease (luminal stenosis $> 50\%$ as assessed by cardiac catheterization, computed tomography angiography, or positive stress nuclear or echocardiography) but had New York Heart Association (NYHA) class I–IV HF based on their clinical symptoms. A diagnosis of DCM was recorded for 38 of the 109 HF patients, but NYHA class was uncertain in one. For analysis, the HCM group included varying etiologies (pressure overload hypertrophy–LVH,

genetic), with 23 cases documented among the 109 HF patients. In addition to the HF patients, there were 24 patients with no HF: 15 of these were 31–83 years who had anterior MI 7 weeks to 16 years prior to CMRS; the other 9 had HCM and were 41–80 years. The patient's clinical diagnoses and NYHA class for HF severity were used as 'ground truth' for teaching and evaluating the accuracy of neural network classifications. The demographics of the study group are summarized in Table 1.

CMR/CMRS methods

One hundred and twenty-four studies (31 healthy, 20–60 years; 94 patients, 22–83 years) were performed at 1.5 T using the 'four-angle saturation-transfer' (FAST) method [19] for measuring k_f , and a combined internal (water) and external (phosphate) referencing method [15, 16] for measuring metabolite concentrations in the anterior myocardium and septum as identified by MRI [6]. Fifty-four subjects (14 healthy, 22–65 years, $p = 0.99$ vs 1.5 T; 40 HF patients, 20–69 years, $p = 0.46$ vs 1.5 T HF patients) were studied at 3 T using cardiovascular magnetic resonance (CMR)-guided 'triple repetition time saturation-transfer' (TRiST) [20] for measuring k_f , and an external (phosphate) reference for the anterior myocardial metabolite concentrations [18]. One-dimensional chemical shift image localization was used with the same 1-cm spatial resolution for all acquisitions in each ^{31}P CMRS protocol, with exam times ranging from 45 to 75 min. Figure 1 shows examples of raw CMR and ^{31}P CMRS saturation transfer recordings at 1.5 T from healthy, DCM, anterior MI and hypertrophy subjects.

Cardiac [PCr], [ATP], and k_f were measured from CMRS peaks with signal-to-noise ratios (SNR) > 5 [22, 23] in the frequency domain, using simplex Gaussian fitting at 1.5 T and 'Circle fit' at 3 T. These analysis methods have

been shown to produce indistinguishable results [28]. Each metabolic data set was comprised of an average of values from 2 to 4 sections of the anterior myocardial wall of each subject, corrected for partial saturation and spill-over irradiation of k_f according to the latest published accounts [21–23]. No CMRS or clinical data were altered, filtered, reanalyzed, or corrected for this study, and no existing data sets were knowingly excluded. All data were treated as fully-corrected measurements of equal weighting and validity. The acquisition and data analysis protocols used for these studies were detailed in the original reports [6, 11, 21–23, 26, 27].

Neural-network analysis

The machine learning and neural-network tools in MATLAB 2017A (Mathworks Inc., Natick, Massachusetts, USA) were used throughout. The MATLAB 'Classification Learner App' was first used to test whether non-neural network machine-learning approaches could be used to differentiate pooled HF data from pooled healthy and non-HF studies. The utility of the cubic 'k nearest neighbor' (KNN) cubic spline fitting; Gaussian 'support vector machine' (SVM); and 'bagged decision tree' ('random forest' bagging, involving the random artificial duplication of data sets) methods were also tested, but their classification accuracy appeared limited to < 70% for each method with some evidence of over-fitting, so these methods were not pursued.

Seeking better performance, we designed a first neural network with two stacked MATLAB 'auto-encoder' layers, followed by a 'softmax' layer to classify the four-dimensional metabolic data (ATP, PCr, k_f , and CK Flux) by disease type and HF severity. The auto-encoders reduce data dimensionality via compression (encoding)

Table 1 Numbers of subjects (n) studied by noninvasive 1.5 T and 3 T ^{31}P CMRS between 1999 and 2011 and their age (mean \pm SD), gender, and clinical assessment at the time of CMRS

	1.5 T (n = 124)	3 T (n = 54)	HCM	DCM	NYHA class I	NYHA class II	NYHA class III	NYHA class IV
Control	31	14						
Age (yr)	40 \pm 8	40 \pm 18						
M/F	20/11	7/7						
HF-NICM	69	40	23	38	17	56	32	4
Age (yr)	48 \pm 11	46 \pm 14	48 \pm 13	49 \pm 12	44 \pm 13	48 \pm 12	47 \pm 12	44 \pm 5
M/F	43/26	16/26	11/12	29/9	10/7	27/29	20/12	2/2
HCM-no HF	9		9					
Age (yr)	59 \pm 13		59 \pm 13					
M/F	5/4		4/5					
Anterior MI-no HF	15							
Age (yr)	57 \pm 16							
M/F	12/3							

HF Heart failure, NICM nonischemic cardiomyopathy; M/F Male/Female, HCM Hypertrophic cardiomyopathy, MI Myocardial infarction, NYHA New York Heart Association (clinical classification for HF)

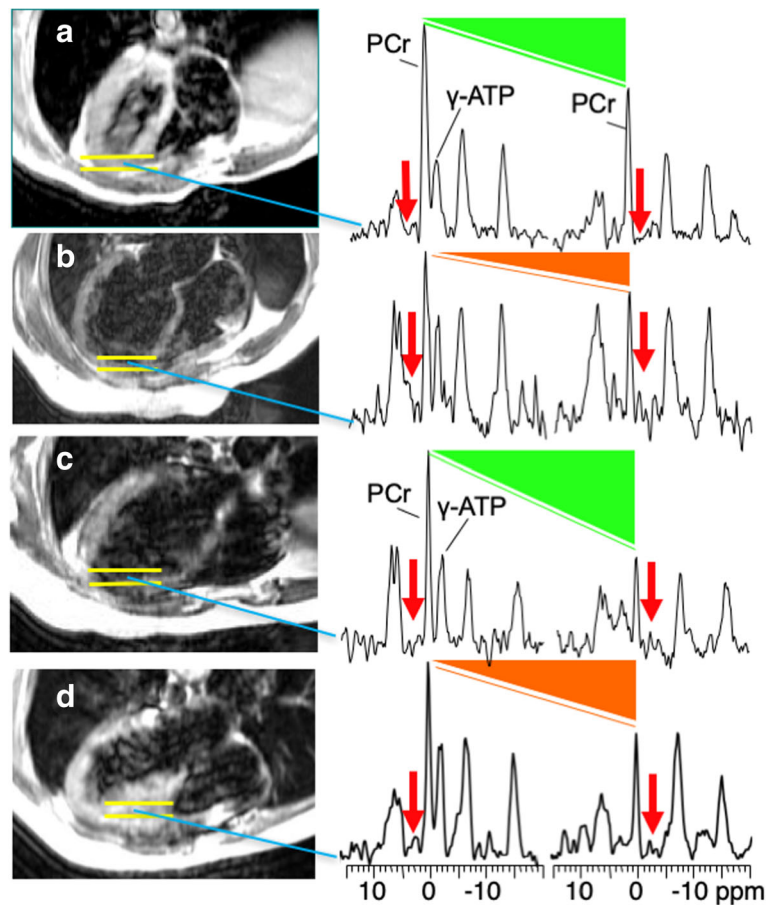
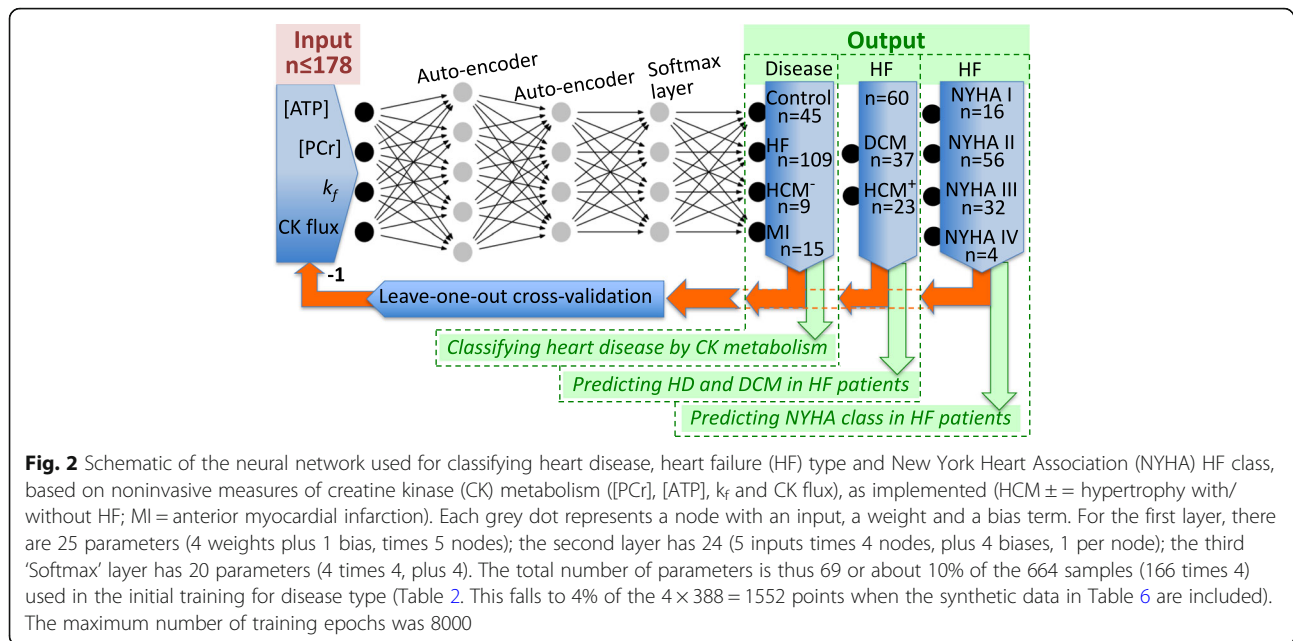


Fig. 1 Examples of axial scout CMR (left) and two of the four FAST saturation transfer spectra (right) [19] from a single section in the anterior myocardium (yellow bars on CMR, left panel) of a healthy subject [6] (a); a patient with a dilated cardiomyopathy (DCM) and NYHA Class III heart failure (HF) [6] (b); one with anterior myocardial infarction (MI) and no HF [21] (c); and one with hypertrophic cardiomyopathy (HCM) and NYHA Class I HF [22] (d). Red arrows on the spectra denote the location of saturating irradiation which causes changes in the phosphocreatine (PCr) peak proportional to the creatine kinase (CK) pseudo first-order reaction rate, k_f (green triangles, no-HF; orange triangles, HF). In the absence of saturating radiation the PCr and adenosine triphosphate (ATP) signals are proportional to the concentrations, [PCr] and [ATP] which are measured with a reference concentration. For these subjects the ^{31}P CK metabolic parameters of [PCr] in $\mu\text{mol/g}$, [ATP] in $\mu\text{mol/g}$, CK flux in $\mu\text{mol/g/sec}$ and k_f in sec^{-1} were (12, 4.3, 3.3, 0.27), (8.5, 6.9, 2.5, 0.29), (6.9, 4.2, 2.7, 0.39), and (6.1, 4.3, 2.8, 0.47) respectively

and decompression (decoding) functions that map the input data onto itself (hence ‘auto’) through back-propagation. The encoder’s output is the compressed representation of the input, which is either fed to the second auto-encoder layer, or to the network’s final output. Two auto-encoders were chosen to provide flexibility in dimensional reduction, while minimizing over-fitting that can result from either the rigidity of using a single auto-encoder, or the complexity introduced by more layers. Figure 2 shows a schematic of the model network as implemented.

The network was ‘trained’ by iteratively adjusting the weights and bias parameters associated with the individual network nodes using L2-regularization (performed internally within the MATLAB App ‘train’ function; L2 regularization parameter = 0.0025), and ‘leave-one-out’ cross-validation. The initial training utilized 166 of the data sets, omitting 12 patients who had both HCM and

HF which would interfere with the fellow class of HCM without HF. Leave-one-out cross-validation can be considered equivalent to 166-fold cross-validation, as it is repeated 166 times with a different set of patient data omitted from each run to determine the corresponding accuracy and to adjust the hyper-parameters. The fluctuation in accuracy across the set of cross-validation outputs was used to fine-tune the hyper-parameters to achieve uniform (or ‘balanced’) accuracy of about 70% across all target classes. The network’s performance was then determined using all of the training data, with the results presented as a ‘confusion-matrix’, as ‘receiver operating curves’ (ROCs) which plot sensitivity versus (1-specificity) for the neural network classifications, and as ‘area-under-the-curve’ (AUC) analyses. The classification of data omitted from the training sets is reported separately.



Second, in order to determine which, if any, metabolic parameter(s) contributed most to disease classification and the potential impact of using fewer variables on prediction, we investigated the effect of the systematic elimination of input parameters from the network. After each omission, the network was redesigned to accommodate the reduced number of input parameters, while keeping the number of layers the same. The redesigned network was re-trained to optimize the hyper-parameters for the network using leave-one-out cross-validation, and the final performance analyzed to determine the change in accuracy and AUC.

Third, we tested whether a CK metabolism-based neural network could distinguish subgroups of the HF patient group who were also documented with DCM ($n = 37$, with the case of uncertain NYHA class excluded from the training set) and HCM ($n = 23$, now including the 12 HCM cases with HF omitted from the initial analysis). This (third) network was trained and optimized using the same methods described above.

Fourth, we tested whether CK metabolism could distinguish the degree of clinical HF severity as indexed by NYHA class among the subgroup of 109 HF patients. Here we used a (fourth) two-stage serial network because the closer proximity of the NYHA class II and III metabolic data (vs. the NYHA class I and class IV data) compromised network balance when all four classes were fitted simultaneously. Thus the HF data were first classified into three classes: NYHA class I; NYHA classes II plus III combined; and NYHA class IV. The output of the combined NYHA classes II and III of the first network stage, were input to a second neural network classifier to distinguish the two. Omitting the

uncertain NYHA case, this study included 108 of the 109 HF cases.

Confounding factors

Because the distribution of patients studied at 3 T differed from those at 1.5 T, we further tested whether the classification of heart disease from CK metabolic data was confounded by the use of the different methodologies rather than their underlying metabolic characteristics. For this, the performance of the first network that was trained to classify the control, HF and non-HF patients using the combined 1.5 and 3 T data, was tested separately on the 1.5 T and the 3 T data sets without retraining. Next, we retrained and re-optimized the network using the 1.5 T data alone, and re-tested its performance on the 1.5 T data and on the pooled 1.5 T plus 3 T data.

To test the effects of unbalanced sampling of both disease type and HF NYHA class on neural network performance without discarding any patient data, we used the 'synthetic minority over-sampling technique' (SMOTE) to augment the original class sets with data randomly interpolated between nearest neighbors in parametric space [29, 30]. Each synthetic point was randomly interpolated from one of five nearest-neighbor pairs of each of the three independent parameters, [PCr], [ATP] and k_f . CK flux, which is a dependent parameter, was derived from the product of the new [PCr] and k_f data. Points were added until the disease class-size matched that of the majority training class (HF, $n = 97$). Points in excess of the targeted class size that arose from the multi-fold interpolation were omitted as outliers, as assessed by the original neural network, analogous to the 'generative adversarial neural network technique' [31]. Similarly, NYHA

class size was balanced by SMOTE interpolation to match that of the majority Class II HF ($n = 56$), after discarding the Class IV data ($n = 4$) as too limited for this purpose. Given the larger data sets generated by this process, the neural networks for disease type and HF severity were retrained with the synthetically-balanced data sets using '5-10 fold cross-validation' (that is, an 80–90% training versus 20–10% testing data split) instead of leave-one-out cross-validation. The performance was then re-assessed. Finally, the performance of the balanced retrained network was tested with the original data; and the network trained with the original data was tested with the augmented HF data.

Results

Table 2 presents the confusion-matrix documenting the accuracy of the first network for classifying healthy and heart disease patients. The results show an overall accuracy of 84% for discriminating heart disease solely on the basis of CK metabolism alone. The network correctly distinguished 90% of the HF data, and about

Table 2 Confusion matrix for the prediction of heart disease for the (first) neural network using the four CK metabolic parameters, [PCr], [ATP], k_f , and CK flux). The numbers at the top of the boxes are the number of cases classified correctly (green boxes), or misclassified (orange boxes): their sum is the total classified (166 = 178–12 HCM patients who had HF and were omitted from the training set). The percentages of the total are listed below the numbers (eg, for HF, $87/166 = 52\%$ of the sample). At bottom, the % correct (true positive rate, green bold font) and incorrect (false negative rate, red bold font) are summarized for each output disease class (HF, HCM, MI and Control). For example, in the first column 87 HF cases were correctly classified. An additional 10 HF cases were incorrectly assigned as non-HF (1 as MI and 9 as Controls). Thus $87/97 = 90\%$ are correct and 10% are incorrect. The percentages in the right-hand column (positive predictive value, green font; false discovery rate, red font) include subjects from all classes. For example, a total of 102 cases were classified by the network as HF. This comprises 87 true HF cases, plus 1 HCM, 5 MI and 9 Controls all misclassified as HF. Thus, $87/102 = 85\%$ are correct. Summing the diagonal, $140/166 = 84\%$ of cases are correctly classified overall

HF	87 52%	1 0.6%	5 3%	9 5.4%	85% 15%
HCM (no HF)	0 0%	7 4%	0 0%	0 0%	100% 0%
MI (no HF)	1 0.6%	0 0%	10 6%	0 0%	91% 9%
Control	9 5%	1 0.6%	0 0%	36 22%	78% 22%
Correct	90%	78%	67%	80%	84%
Incorrect	10%	22%	33%	20%	16%
Target class:	HF	HCM (no HF)	MI (no HF)	Control	

80% of the other classes except for MI, which was difficult to differentiate from HF (accuracy $\sim 70\%$). It predicted the primary diagnosis of HF in the 12 HCM cases that were omitted from the training set (and Table 2), with 100% accuracy.

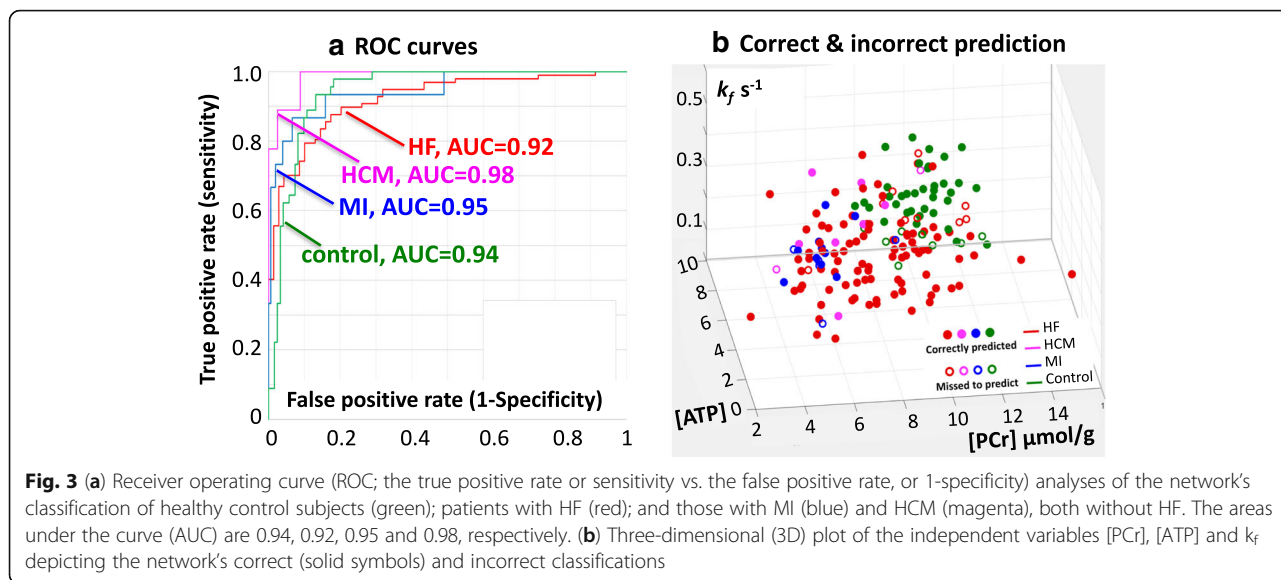
Figure 3(a) illustrates the results of the ROC analyses of the (first) network's classification of each disease using the four CK metabolic parameters. The AUCs are all greater than 90%. Figure 3(b) is a three-dimensional (3D) plot of the independent variables [PCr], [ATP] and k_f depicting the network's true and false classifications. A 3D video plot (see Additional file 1) shows that the fitted volumes associated with each output disease type are intertwined on the [PCr]–[ATP] axis (diagonal), indicating that the PCr/ATP ratio, by itself, is a poor discriminator of disease type.

The effect of successively eliminating in turn, each CK metabolic parameter and re-tuning the network in the second set of studies is shown in Fig. 4. All variables contributed to discrimination, but [ATP] contributed least to overall accuracy. The effect of omitting metabolic variables on the ROC–AUC analysis, for each disease type (Fig. 5) shows that: (i) successively removing variables reduces the AUC; (ii) patients with HCM but no HF were hardest to classify; and (iii) [ATP] provided the least added-value as a discriminator, as evidenced by the fact that including [ATP] instead of any other variable produced the greatest loss in AUC.

Table 3 shows the confusion matrix for the (third) network designed to distinguish HF patients documented with DCM from those with HCM, based just on their CK metabolism. Again an accuracy of about 80% was achieved. The network predicted the one DCM subject whose NYHA class was uncertain as class II HF (presently not confirmable), with a 50/50 chance of being either DCM or HCM.

Table 4 presents the confusion matrix results for classifying NYHA HF severity using the (fourth) CK metabolism-based neural network classifier. The overall accuracy for discriminating NYHA HF class on the basis of CK metabolism alone was 86%, with the caveat that the number of NYHA class IV patients is very limited.

When all the 3 T data are omitted from the first neural network that was used for Table 2 in order to assess whether the results are confounded by differences in the 1.5 T and 3 T methodologies, the accuracy of the resulting confusion matrix (Table 5a) improved slightly to 86% (vs 84% in Table 2). Omitting the 1.5 T data instead of the 3 T data yielded a similar 82% accuracy using this same un-retuned (first) network (Table 5b). Retraining and optimizing this network for just the 1.5 T data alone, improved the overall accuracy to 90% (Table 5c). This may be due to some systematic differences (including variability in scatter)



between the 1.5 T and 3 T data and/or differences in the sub-groups sampled.

Rebalancing the network used for classifying disease with randomly interpolated data to equalize class size to that of HF, resulted in the same overall accuracy of 84% (Table 6a). This reflects an improvement by about 10% in the accuracy of detecting the formerly lesser-sampled LVH, MI and control groups, at the expense of HF patients whose accuracy declined by 14%. Similarly, rebalancing the network for distinguishing HF severity by augmenting the NYHA Classes I and III data to match the Class II data pool-size, resulted in essentially the same overall accuracy (85% vs. 86% in Table 4), with performance improving for Classes I and III, at the expense of Class II (Table 6b). Applying the balanced retrained network to the original data reduced overall performance a little to 78%, mainly to the detriment of the small HCM non-HF group (accuracy, 67%). Finally, when the network that was trained with the original data from Table 2, was plied with the entire set of

augmented HF data ($n = 168$), its overall accuracy for detecting HF was 85%.

Discussion

This is the first study to apply neural network techniques to test whether cardiac CK energy metabolism can differentiate the type of myocardial disease and the severity of disease in individual patients. The source data comprised all of the known complete sets of individual cardiac [PCr], [ATP], k_f and CK flux data from healthy subjects and heart disease patients extant in the world that we know of [1, 6, 11, 21–23, 26, 27]. The results show that a three-layer neural-network can differentiate healthy, HF, non-HF MI and HCM cardiac diseases with an accuracy of 84% and an AUC of > 90% for each category, based on their CK metabolism alone (Table 2, Fig. 3). Among HF patients a neural network was able to differentiate DCM and HCM (Table 3), as well as the

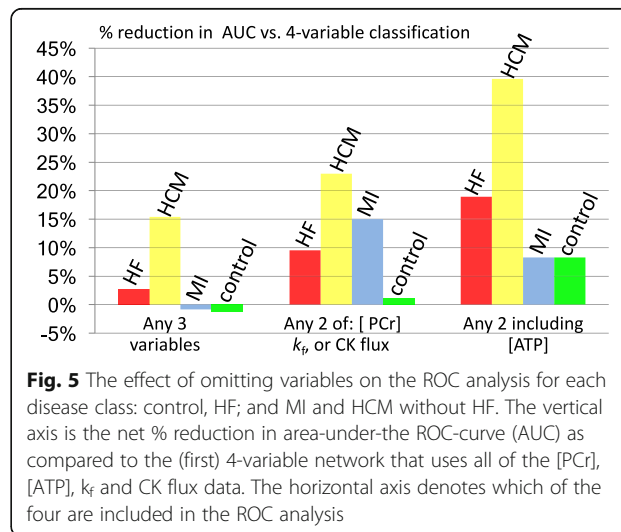
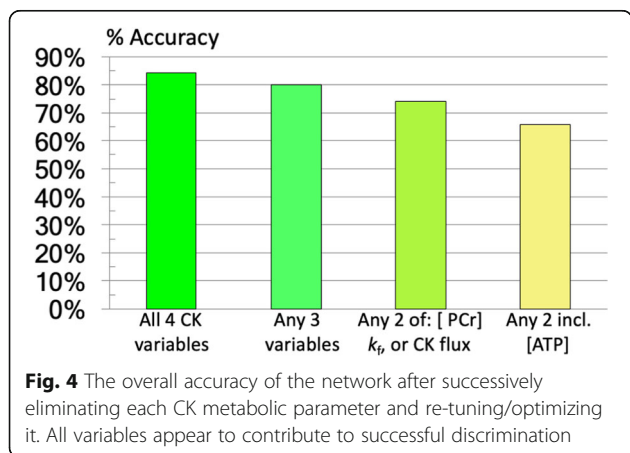


Table 3 Confusion matrix for predicting HCM ($n = 23$ cases) and DCM ($n = 37$, excluding one with uncertain NYHA class) among HF patients with NYHA class $\geq I$ using a neural network based on the four CK metabolic parameters, [PCr], [ATP], k_f , and CK flux. The numbers at the top of the boxes are the number of cases classified correctly (green boxes) or misclassified (orange boxes), with the percentages of the total listed below. The % correct (green, bold font) and incorrect (red, bold font) are summarized for each output class (see Table 1 caption for interpretation)

DCM (HF)	29 48%	5 8%	85% 15%
HCM (HF)	8 13%	18 30%	69% 31%
Correct	78%	78%	78%
Incorrect	22%	22%	22%
Target class:	DCM (HF)	HCM (HF)	

severity of HF as indexed by NYHA class (Table 4), both with about 80% or higher overall accuracy. This suggests that CK metabolic data acquired noninvasively with resting-state cardiac ^{31}P CMRS, could be used with a neural network classifier as a metabolic biomarker to stratify disease type and severity with clinically-relevant accuracy. Moreover, from a mechanistic standpoint, the fact that a neural network can classify disease based on CK metabolism alone, means that there must exist heretofore unidentified differences in CK metabolism that are specific to

Table 4 Confusion matrix for the prediction of NYHA class I-IV for HF among HF subjects using a two-stage neural network and the four CK metabolic parameters [PCr], [ATP], k_f , and CK flux as input. The numbers at the top of the boxes are the number of cases classified correctly (green boxes), or misclassified (orange boxes), totaling 108 cases excluding the one case with uncertain NYHA class (See text). The percentages of the total are listed below the numbers. The % correct (green, bold font) and incorrect (red, bold font) are summarized for each output NYHA class (see Table 1 caption to guide interpretation)

NYHA I	12 11%	0 0%	0 0%	0 0%	100% 0%
NYHA II	4 4%	54 50%	9 8%	0 0%	81% 19%
NYHA III	0 0%	2 2%	23 21%	0 0%	92% 8%
NYHA IV	0 0%	0 0%	0 0%	4 4%	100% 0%
Correct	75%	96%	72%	100%	86%
Incorrect	25%	4%	28%	0%	14%
Target class:	NYHA I	NYHA II	NYHA III	NYHA IV	

Table 5 Confusion matrices for the prediction of heart disease for the neural network used in Table 2, applied to: (a) the 1.5 T data alone ($n = 112$ omitting 12 interfering cases with both HD and HF); and (b) the 3 T data alone ($n = 54$ cases). (c) Confusion matrix after retuning and optimization using only the 1.5 T data. (NA = not applicable due to absence of 3 T data for that class)

(a)	HF	51 46%	1 0.9%	5 4.5%	3 2.7%	85% 15%
	HCM (no HF)	0 0%	7 6.3%	0 0%	0 0%	100% 0%
	MI (no HF)	1 0.9%	0 0%	10 9%	0 0%	91% 9%
	Control	5 4.5%	1 0.9%	0 0%	28 25%	82% 28%
	Correct	90%	78%	67%	90%	86%
	Incorrect	10%	22%	33%	10%	14%
(b)	HF	36 67%	0 0%	0 0%	6 11%	86% 14%
	HCM (no HF)	0 0%	0 0%	0 0%	0 0%	NA NA
	MI (no HF)	0 0%	0 0%	0 0%	0 0%	NA NA
	Control	4 7.4%	0 0%	0 0%	8 15%	82% 28%
	Correct	90%	NA	NA	57%	82%
	Incorrect	10%	NA	NA	43%	18%
(c)	HF	52 46%	3 2.7%	2 1.8%	0 0%	91% 9%
	HCM (no HF)	1 0.9%	6 5.4%	1 0.9%	0 0%	75% 25%
	MI (no HF)	2 1.8%	0 0%	12 11%	0 0%	86% 14%
	Control	2 1.8%	0 0%	0 0%	31 28%	94% 6.1%
	Correct	92%	67%	80%	100%	90%
	Incorrect	9%	33%	20%	0%	10%
	Target class:	HF	HCM (no HF)	MI (no HF)	Control	

disease type and severity which enable the network to do so.

Indeed there was no one unique disease-specific CK-metabolic defect, as evidenced by the incremental effect of successively removing each parameter and retuning the network (Fig. 4). Instead, nuanced variations in multiple parameters permitted accurate neural network discrimination. The elimination analysis showed that [PCr] and k_f were the first and second most important input parameters for classifying heart disease in the neural network (Figs. 4 and 5). The pair, [PCr] and k_f , was better than CK flux alone, but using all three, namely CK flux, k_f , and [PCr], was better still. This is unexpected because CK flux is not independent, being the product of k_f and [PCr] as noted earlier. In order for CK flux to add such value as a classifier compared to its component factors, k_f and [PCr] must vary in a co-dependent way that is exacerbated by disease. This is consistent with the observation that CK flux independently predicted cardiac

Table 6 Confusion matrices for predicting heart disease and HF severity based on neural networks retrained with the augmented synthetic [PCr], [ATP] and k_f data to balance class size. (a) Results from the network as in Table 2, retrained with 388 cases comprised of groups of 97 with HF, and 97 synthetically-augmented cases of HCM, MI and control subjects. (b) Results from the network as in Table 4, retrained with 168 cases of HF comprised of 56 patients with NYHA Class II severity and two synthetically-augmented groups of 56 each, with NYHA Class I and III

(a)					
HF	74 19%	7 2%	5 1%	6 1.5%	80% 20%
HCM (no HF)	10 2.6%	88 23%	16 4%	4 1%	75% 25%
MI (no HF)	4 1%	1 0.3%	76 20%	0 0%	94% 6%
Control	9 2.3%	1 0.3%	0 0%	87 22%	90% 10%
Correct Incorrect	76% 24%	91% 9%	78% 22%	90% 10%	84% 16%
Target class:	HF	HCM (no HF)	MI (no HF)	Control	
(b)					
NYHA I	52 31%	4 2%	0 0%	93% 7%	
NYHA II	1 0.6%	45 27%	10 6%	80% 20%	
NYHA III	3 1.8%	7 4%	46 27%	82% 18%	
Correct Incorrect	93% 7%	80% 20%	82% 18%	85% 15%	
Target class:	NYHA I	NYHA II	NYHA III		

events whereas individual k_f , [PCr] (and PCr/ATP) measures alone, did not [22]. [ATP] was the least important parameter for discriminating disease, an observation likely reflecting the tight control of [ATP] that is required to preserve cellular viability. With [ATP] as the 'last to go', the CK reaction, oxidative phosphorylation, glycolysis and other metabolic pathways apparently make whatever adjustments that are needed to preserve [ATP] and prevent demise.

Limitations

The main limitation of this work is the number of studies, particularly the NYHA class IV HF, and non-HF HCM and MI sub-categories. Although all extant complete data sets were included, they derive from research studies that were not designed for the specific investigation reported herein. In that respect the study might be considered as blinded compared to the prospective acquisition of data to test a preconceived hypothesis. Unfortunately, the availability of new data to

test the potential of this cardiac ^{31}P CMRS quartet to serve as a metabolic biomarker for clinical disease, is presently limited to the extent that cardiac ^{31}P MRS is not yet a clinical procedure.

The synthetic augmentation of metabolic data by 2.3-fold for disease type and by 1.7-fold for HF severity using the SMOTE method, provided results that were consistent with the smaller original data sets. This suggests that the overall level of performance realized here should be reasonably representative of what is achievable with the addition of ^{31}P CMRS data of comparable quality from future studies. Nevertheless, the primary purpose of data augmentation was to balance class-size to avoid over-fitting and/or bias of the neural network towards a majority class. Indeed, its overall effect was to smooth out performance differences between the majority and lesser-sampled classes (Table 6 vs. Tables 2 and 4), which was the intended purpose [29]. There is no substitute for original data, especially when the underlying functional inter-relationship between metabolic parameters, disease type and severity is not known. That the successive dropping of metabolic parameters degraded predictive accuracy (Fig. 4), means that [PCr], [ATP], k_f and CK flux have to be treated as a parametric quartet. Thus in general, network performance might not be expected to survive simulations that involve independent randomization of the component parameters, at least until the nature of their mechanistic link is better understood. Disease heterogeneity is a potentially confounding factor as hypertrophy, DCM and MI are common co-morbidities that underlie HF. They would therefore be expected to confound the neural network's ability to distinguish them. In that sense, the accuracy with which the network was able to discriminate disease type as demonstrated here, is quite surprising. Also surprising was its ability to differentiate NYHA class, which involves a degree of subjectivity in clinical assessment. Indeed all of the network's misclassification errors involved sliding into an immediately adjacent NYHA class (Table 4).

Other potential confounding variables were age and gender differences, and the use of different 1.5 and 3 T methodologies among sub-groups of the study population. As can be seen in Table 1, the former are fairly minor with the ages of control and HF patients being around 40–50 years. Although the MI and non-HF HCM patients were older, these patients had normal k_f and [ATP] when compared with their original study controls [11, 21]. Nevertheless, a trend for reduced metabolite concentrations (and probably PCr/ATP) with age over 40 [1] has been reported elsewhere and should be noted. Regarding methodological differences between the 1.5 and 3 T CMRS protocols, the re-analysis of separated 1.5 T and 3 T data should eliminate any such

confounding effects (Table 5). Indeed, the ability of the neural network to classify disease based on CK metabolism alone was unimpaired for those diseases for which CK metabolic data were available from both 1.5 T and 3 T methodologies. These results can only further support the primary conclusion that a neural network can differentiate individual disease type and severity based on CK metabolism alone.

Conclusion and clinical implications

Neural network analysis employing only cardiac CK metabolic parameters can accurately classify cardiac disease and severity, suggesting a role for noninvasive quantitative ^{31}P CMRS as a metabolic biomarker. In HF, the degree of compromise in CK energy supply as identified by the neural network, predicted the clinical severity of HF symptoms in individual patients. The ability to accurately classify disease based on CK metabolism suggests that altered CK metabolism is an intrinsic and nuanced factor in many common heart diseases. If the relationship between compromised CK metabolism and clinical severity is causal, then therapies that increase CK energy supply could well reduce symptoms and support a role for the CK reaction as a potentially important new therapy target [22, 26].

Additional file

Additional file 1: CK metabolism disease classifier and fitting volumes in 3D. (MP4 3733 kb)

Abbreviations

3D: Three dimensional; ADP: Adenosine diphosphate; ATP: Adenosine triphosphate; AUC: Area-under-the-curve; CK: Creatine kinase; CMR: Cardiovascular magnetic resonance; CMRS: Cardiovascular magnetic resonance spectroscopy; Cr: Unphosphorylated creatine; DCM: Dilated cardiomyopathy; FAST: Four-angle saturation transfer; HCM: Hypertrophic cardiomyopathy; HF: Heart failure; KNN: K nearest neighbor; LV: Left ventricle/left ventricular; LVH: Left ventricular hypertrophy; MI: Myocardial infarction; MRI: Magnetic resonance imaging; NICM: Non-ischemic cardiomyopathy; NYHA: New York Heart Association; PCr: Phosphocreatine; ROC: Receiver operating curve; SD: Standard deviation; SMOTE: Synthetic minority over-sampling technique; SNR: Signal- to -noise ratio; SVM: Support vector machine; T: Tesla (unit); TRIST: Triple repetition time saturation-transfer

Authors' contributions

PAB conceived the study, provided the data pool and wrote the manuscript. MS performed the neural network analyses that created Tables 2, 3, 4, and 5 and Figs. 2, 3 and 4 and the initial draft of the Methods section. PAB and RGW were responsible for the collection and analysis of the original raw data. All authors read, corrected and edited the manuscript. All authors read and approved the final manuscript.

Funding

Supported by: the Russell H Morgan Professorship at Johns Hopkins University (P.A.B., M.S.), the Newton Abraham Visiting Professorship at Lincoln College, Oxford University (P.A.B.); and N.I.H. grant R01HL61912 (R.G.W., P.A.B.)

Availability of data and materials

Data generated during this study are included in this published article, the supplementary file, [ClinicalTrials.gov](https://clinicaltrials.gov) (Identifier, NCT00181259), and the cited references.

Ethics approval and consent to participate

The original data were acquired from studies approved by the Johns Hopkins University Institutional Review Board, wherein all participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Division of MR Research, Department of Radiology, Johns Hopkins School of Medicine, Park Bldg. 310, 600 N Wolfe St, Baltimore, MD 21287, USA.

²Division of Cardiology, Department of Medicine, Johns Hopkins University, School of Medicine, Baltimore, MD, USA.

Received: 5 November 2018 Accepted: 1 July 2019

Published online: 12 August 2019

References

1. Bottomley PA. MRS studies of creatine kinase metabolism in human heart. *eMagRes*. 2016;5:1183–202. <https://doi.org/10.1002/9780470034590.emrstm1488>.
2. Weiss RG, Bottomley PA, Hardy CJ, Gerstenblith G. Regional myocardial metabolism of high-energy phosphates during isometric exercise in patients with coronary artery disease. *N Engl J Med*. 1990;323:1593–600.
3. Hardy C, Weiss RG, Bottomley PA, Gerstenblith G. Altered myocardial high-energy phosphate metabolites in patients with dilated cardiomyopathy. *Am Heart J*. 1991;122:795–801.
4. Neubauer S, Krahe T, Schindler R, Horn M, Hillenbrand H, Entzeroth C, Mader H, Kromer EP, Riegger GAJ, Lackner K, Ertl G. ^{31}P magnetic resonance spectroscopy in dilated cardiomyopathy and coronary artery disease. *Circulation*. 1992;86:1810–8.
5. Neubauer S, Horn M, Cramer M, Harre K, Newell JB, Peters W, Pabst T, Ertl G, Hahn D, Ingwall JS, Kochsiek K. Myocardial phosphocreatine-to-ATP ratio is a predictor of mortality in patients with dilated cardiomyopathy. *Circulation*. 1997;96:2190–6.
6. Weiss RG, Gerstenblith G, Bottomley PA. ATP flux through Creatine kinase in the Normal, stressed, and failing human heart. *Proc Natl Acad Sci U S A*. 2005;102:808–13.
7. Conway MA, Allis J, Ouwerkerk R, Nioka T, Rajagopalan B, Radda GK. Detection of low PCr to ATP ratio in failing hypertrophied myocardium by ^{31}P magnetic resonance spectroscopy. *Lancet*. 1991;338:973–6.
8. Conway MA, Bottomley PA, Ouwerkerk R, Radda GK, Rajagopalan B. Mitral regurgitation: impaired systolic function, eccentric hypertrophy, and increased severity are linked to lower phosphocreatine/ATP ratios in humans. *Circulation*. 1998;97:1716–23.
9. Lamb HJ, Beyerbrach HP, van der Laarse A, et al. Diastolic function in hypertensive heart disease is associated with altered myocardial metabolism. *Circulation*. 1999;99:2261–7.
10. Beer M, Seyfarth T, Sandstede J, Landschutz W, Lipke C, Kostler H, von Kienlin M, Harre K, Hahn D, Neubauer S. Absolute concentrations of high-energy phosphate metabolites in normal, hypertrophied, and failing human myocardium measured noninvasively with ^{31}P -SLOOP magnetic resonance spectroscopy. *J Am Coll Cardiol*. 2002;40:1267–74.
11. Smith CS, Bottomley PA, Schulman SP, Gerstenblith G, Weiss RG. Altered Creatine kinase adenosine triphosphate kinetics in failing hypertrophied human myocardium. *Circulation*. 2006;114:1151–8.
12. Holloway CJ, Cochlin LE, Emmanuel Y, Murray A, Codreanu I, Edwards LM, Szmigielski C, Tyler DJ, Knight NS, Saxby BK, Lambert B, Thompson C, Neubauer S, Clarke K. A high-fat diet impairs cardiac high-energy phosphate metabolism and cognitive function in healthy human subjects. *Am J Clin Nutr*. 2011;93:748–55.
13. Metzler B, Schocke MFH, Steinboeck P, Wolf C, Judmaier W, Lechleitner M, Lukas P, Pachinger O. Decreased high-energy phosphate ratios in the

- myocardium of men with diabetes mellitus type I. *J Cardiovasc Magn Reson*. 2002;4:493–502.
14. Bottomley PA, Weiss RG, Hardy CJ, Baumgartner WA. Myocardial high-energy phosphate metabolism and allograft rejection in patients with heart transplants. *Radiology*. 1991;181:67–75.
 15. Bottomley PA, Hardy CJ, Roemer PB. Phosphate metabolite imaging and concentration measurements in human heart by nuclear magnetic resonance. *Magn Reson Med*. 1990;14:425–34.
 16. Bottomley PA, Atalar E, Weiss RG. Human cardiac high-energy phosphate metabolite concentrations by 1D-resolved NMR spectroscopy. *Magn Reson Med*. 1996;35:664–70.
 17. Yabe T, Mitsunami K, Inubushi T, Kinoshita M. Quantitative measurements of cardiac phosphorus metabolites in coronary artery disease by 31P magnetic resonance spectroscopy. *Circulation*. 1995;92:15–23.
 18. El-Sharkawy AEM, Gabr RE, Schär M, Weiss RG, Bottomley PA. Quantification of human high-energy phosphate metabolite concentrations at 3T with partial volume and sensitivity corrections. *NMR Biomed*. 2013;26:1363–71. <https://doi.org/10.1002/nbm.2961>.
 19. Bottomley PA, Ouwerkerk R, Lee RF, Weiss RG. Four angle saturation transfer (FAST) method for measuring creatine kinase reaction rates in vivo. *Magn Reson Med*. 2002;47:850–63.
 20. Schär M, El-Sharkawy AEM, Weiss RG, Bottomley PA. Triple repetition time saturation transfer (TRIST) 31P spectroscopy for measuring human Creatine kinase reaction kinetics. *Magn Reson Med*. 2010;63:1493–501.
 21. Bottomley PA, Wu KC, Gerstenblith G, Schulman SP, Steinberg A, Weiss RG. Reduced myocardial creatine kinase flux in human myocardial infarction: an in vivo phosphorus magnetic resonance spectroscopy study. *Circulation*. 2009;119:1918–24.
 22. Bottomley PA, Panjrath GS, Lai S, Hirsch GA, Wu K, Najjar SS, Steinberg A, Gerstenblith G, Weiss RG. Metabolic rates of ATP transfer through Creatine kinase (CK flux) predict clinical heart failure events and death. *Sci Transl Med*. 2013;5:215re3, pp1-8. <https://doi.org/10.1126/scitranslmed.3007328>.
 23. Gabr RE, El-Sharkawy AEM, Schär M, Panjrath GS, Gerstenblith G, Weiss RG, Bottomley PA. Cardiac work is related to Creatine kinase energy supply in human heart failure. *J Cardiovasc Magn Reson*. 2018;20:81. <https://doi.org/10.1186/s12968-018-0491-6>.
 24. Ingwall JS, Weiss RG. Is the failing heart energy starved? On using chemical energy to support cardiac function. *Circ Res*. 2004;95:135–45.
 25. Neubauer SS. The failing heart—an engine out of fuel. *N Engl J Med*. 2007;356:1140–51.
 26. Hirsch GA, Bottomley PA, Gerstenblith G, Weiss RG. Allopurinol acutely increases ATP energy delivery in failing human hearts. *J Am Coll Cardiol*. 2012;59:802–8.
 27. Abraham MR, Bottomley PA, Dimaano VL, Pinheiro A, Steinberg A, Traill TA, Abraham TP, Weiss RG. Creatine kinase adenosine triphosphate and phosphocreatine energy supply in a single kindred of patients with hypertrophic cardiomyopathy. *Am J Cardiol*. 2013;112:861–6.
 28. Gabr RE, Ouwerkerk R, Bottomley PA. Quantifying in vivo MR spectra with circles. *J Magn Reson*. 2006;179:152–63.
 29. Chawla NV, Hall LO, Bowyer KW, Kegelmeyer WP. SMOTE: synthetic minority oversampling technique. *J Artif Intell Res*. 2002;16:321–57.
 30. He H, Bai Y, Garcia EA, Li S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In: *Proc Intl J Conf Neural Networks (IJCNN 2008)*; 2008. p. 1322–8. doi: [10.1.1.309.942](https://doi.org/10.1.1.309.942).
 31. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative Adversarial Networks. In: *Proceedings of the international conference on neural information processing systems (NIPS 2014)*; 2014. p. 2672–80.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

