

HeartPerfect: Data Mining in a Large Database of Myocardial Perfusion Scintigraphy

Bernard Hotz¹ and Jean-Philippe Thirion¹

¹ HealthCenter Internet Services, 449 route des crêtes,
06560 Sophia-Antipolis, France
bernard.hotz@healthcenter.com

Abstract. We are presenting a method to obtain diagnosis and prognosis information by searching similar images into a large database of Myocardial Perfusion Scintigraphy (MPS) cases for which diagnosis is known. We are applying similarity measures to cardiac images pre-registered with a template. Our database is composed of 1430 patient cases with associated clinical information. For each new case, we sort all the patients of the database from most to less similar ones and compute a severity criterion, based on a statistical analysis of normal and diseased most similar patients. By varying a threshold on the severity criterion and testing the classification of controlled cases, we have measured the operational characteristic of this test (ROC curves), and shown increased performance in sensitivity and specificity for disease detection with respect to clinicians and to experts in consensus. Through the extension of database to patients' outcome information, we expect to extend this method to prognosis.

1 Introduction

The interpretation of medical images is often based on a comparison with a virtual normal reference image. This reference image can be the average of images of presumed normal subjects. This approach is restrictive because the comparison is made only with respect to a normal model. As there are many possible locations, extensions and degrees for cardiac abnormalities, it is not possible to generate a set of synthetic images representing abnormal cases, which is at the same time compact and extensive. All that we can do is to perform comparison relatively to a very large panel of presumed abnormal subjects, coming from a lot of different sources and supposed to recover the most frequent coronary artery disease patterns. We have designed a prototype, HeartPerfect, which is implementing such concept and relies on a database of 1430 rest and stress Myocardial Perfusion Scintigraphy (MPS) exams.

We first describe the method used to pre-process the data and then give the definition of the similarity measures that we are using. We are then showing one result of most similar patients search using our criteria. We define a severity measure of Coronary Artery Disease (CAD) from most similar patient profiles. At last, we compute the operational characteristic of our method for disease detection by comparing results of severity measure on controlled patients (i.e. patients with known clinical information or angio-verified cases).

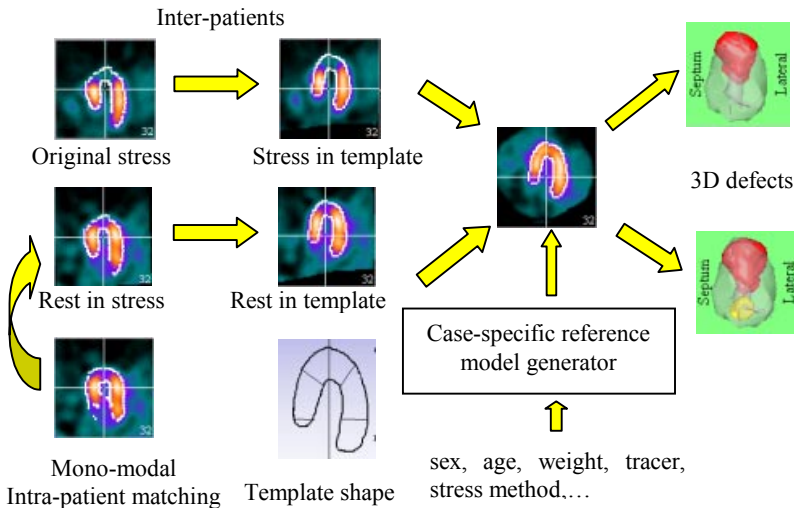


Figure 1: statistical stress and rest perfusion defects computation using CardioMatch® method and a case-specific reference model.

2 Pre-processing of the Database

We are using 1430 stress/rest myocardial perfusion studies (MPS) from five institutions¹ collaborating in a prospective clinical trial to evaluate the performance of the CardioMatch® software (see [2]). All 1430 rest and stress images are registered into a common template geometry using an elastic spline 3D transformation to eliminate size and shape differences (see [1]). After this transformation stress or rest 3D images can be compared between each other at a voxel to voxel basis. The intensities in the image are normalized with respect to a common reference image presenting a normal perfusion, by using a linear regression and outliers rejection. A case specific reference model (CSRM) depending on a database of 100 normal cases as well as on the particular clinical parameters of that case is computed using multi-linear regression. It allows generating an infinity of adapted models without storing a too large amount of data. Parameters such as patient sex, weight and age, stress tracer and stress method are used to compute case-specific normal average and standard deviation images, which are then used to determine statistically significant 3D stress defects (see [3], as well as Figure 1). All cases of the database are pre-processed, once for all, using this method.

¹ Oakland, Santa Clara, South San Francisco, Stanford University and Walnut Creek, Kaiser trial.

3 Similarity Criteria

We have defined three possible ways to obtain lists of similar profiles, with respect to:

1. similar regression parameters.
2. similar stress perfusion images.
3. similar stress perfusion defects.

For a given case, we generate the list of subjects presenting similar clinical parameters. Such parameter-based similarity measure can be used to restrict further exploration to only a subset of the database better suited to the patient's case. This is especially useful when large numbers of subjects are used, which is not our case here, as we have "only" about 1500 cases.

We started with a very simple image-based similarity criterion, which is linear correlation:

$$Linear\ correlation = \frac{\sum_{myocardium} (i_1 - \bar{i}_1)(i_2 - \bar{i}_2)}{\sqrt{\sum_{myocardium} (i_1 - \bar{i}_1)^2 \sum_{myocardium} (i_2 - \bar{i}_2)^2}}$$

More evolved similarity measures can be found in [6] or [4] for brain applications or [7], [8] for non-medical applications. However, the advantage of this criterion is that it is independent of intensity normalization, which reduces one possible cause of variability. It is also relatively fast to be computed: 2'' on a 400MHz PC for a comparison with the 1430 cases of the database.

We have also developed a defects-based similarity criterion, being the overlap percentage of the defects between the reference patient case and the case to compare. The overlap is the number of voxels significantly abnormal in both cases divided by the number of voxels significantly abnormal in at least one case:

$$Defects\ overlap = \frac{card\{v \mid v \in (defects_1 \cap defects_2)\}}{card\{v \mid v \in (defects_1 \cup defects_2)\}}$$

4 Most Similar Patients Search for a Typical Pathological Case

To illustrate our method, we are presenting the 4 most image-based and defects-based similar patients obtained for a typical pathological case (see Figure 2 and Figure 3 respectively).

4.1 Image-Based Similar Cases

In Figure 2 (right column) we present also the average image obtained from all cases whose correlation level is higher than 0.8 (16 cases in this particular example), each profile being weighted by a function of its correlation coefficient. Although this image is very smooth, we can clearly see hypo-perfusion in the infero-lateral location similar to the studied patient’s case. Such image can be used for visual quality control to verify that the cases, which are found, are indeed similar.

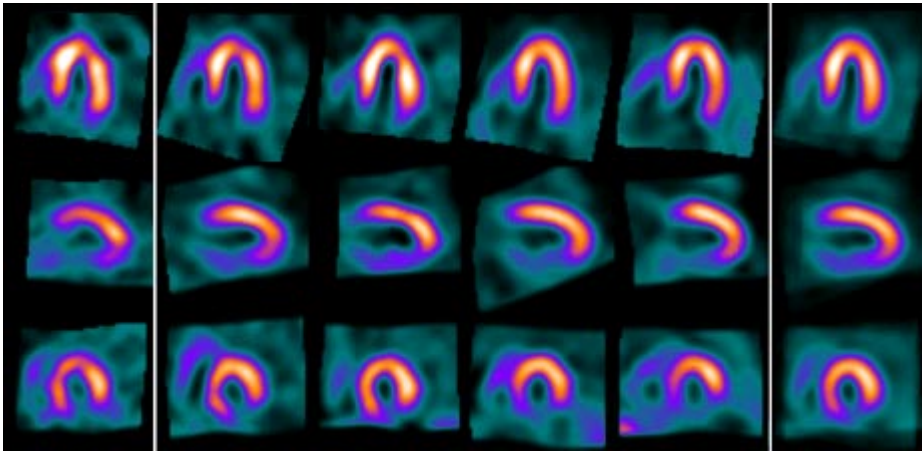


Figure 2: VLA, SA and HLA slices of stress images of the studied case (left). Middle are the 4 most similar images (from most to less similar). Right is the average image for correlation > 0.8.

	Correl.	Sex	Age	Weight	Tracer	Stress	Risk	History	Clini- cians
Case	1.000	M	63	90.0	Thal	exercise	0.943	CABG	Fixed
Sim.1	0.875	M	64	95.0	Tetro	exercise	0.082	CABG	Mixed
Sim.2	0.874	M	61	81.8	Mibi	persantine	0.123	MI	Mixed
Sim.3	0.869	M	76	72.7	Tetro	exercise	0.791	CABG	Fixed
Sim.4	0.848	M	78	68.2	Tetro	persantine	0.943	MI	Ischemic

Table 1: clinical parameters associated with the similar images displayed in Figure 2. Stratification corresponds to Diamond and Forrester (D&F) pre-test probability (see [5]). CABG stands for Coronary Artery Bypass Graft, MI stands for Myocardial Infarction.

We are also giving the clinical information associated to these cases (see Table 1). We will see later on that such information, measured from a larger set of similar patients, can allow us to infer diagnosis and prognosis information.

4.2 Defects-Based Similar Cases

Such measurement makes sense only for patients presenting large perfusion defects (size > 10% of the total myocardium volume, i.e. diseased cases). Results are presented in Figure 3. One interesting finding is that defects computed on the average image obtained with the image-based similarity measure (Figure 2, right column) have generally a larger defects overlap than all other images (Figure 3, right column).

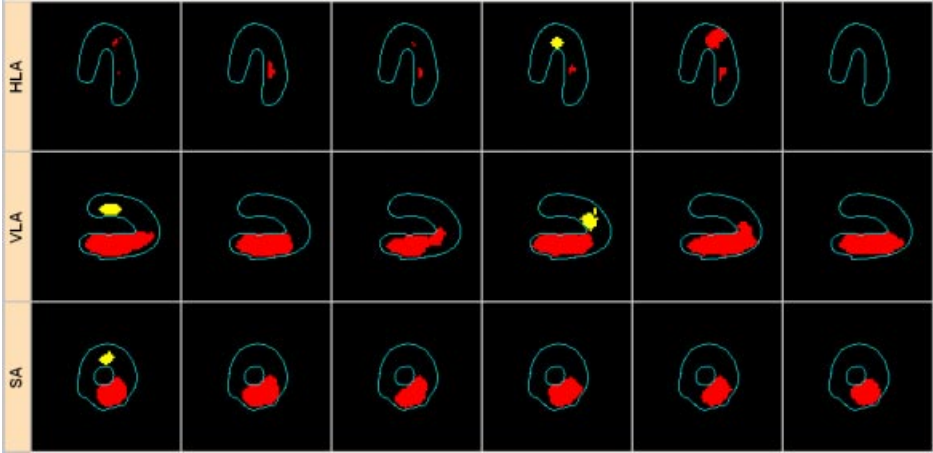


Figure 3: HLA, VLA and SA slices of stress perfusion defects images of the considered case (left), the 4 most similar defects images, from most to less similar (middle) and the defects image which corresponds to average image of Figure 2.

	Over-lap	Sex	Age	Weight	Tracer	S method	Risk	History	Clini-cians
Case	100%	M	63	90.0	Thal	exercise	0.943	CABG	Fixed
Sim.1	68.7%	M	76	72.7	Tetro	exercise	0.791	CABG	Fixed
Sim.2	64.0%	M	72	97.3	Tetro	exercise	0.609	Blank	Ische-mic
Sim.3	63.7%	M	54	77.3	Tetro	exercise	0.998	CABG	Ische-mic
Sim.4	62.8%	M	60	86.8	Tetro	adeno-sine	0.943	ANGIO +	Mixed

Table 2: parameters of similar defect images displayed in Figure 3

5 CAD Severity Measure Based on Similar Patients’ Statistics

The main interest for building lists of similar profiles is to compile clinical information available on similar patients to extrapolate a risk of having a CAD for the considered case. In our database, each MPS was provided with the following information concerning CAD: *CAD risk stratification*, according to both Diamond/Forrester and

Framingham methods. **"Second classification"** according to known history of CAD (10 possible values: angiography (negative, positive or minor), no history of CAD, transplant, evaluation of pre-operative risk, myocardial infarction, coronary artery bypass graft or percutaneous angioplasty, history of mediastinal radiation, percutaneous myocardial reperfusion). **Clinical interpretation** of MPS studies from Nuclear Physicians. **Quantitative results of CardioMatch®** defects analysis (location, relative size and degree of severity of abnormalities as a fraction of the myocardial mass).

In our experiments, we have used only information 2 and 3 to measure the operational characteristics of our new test. Second classification (Class2) is used to declare a case as "positive" (i.e. diseased) if it has a positive angiography, a Myocardial Infarction or a D&F pre-test probability larger than 0.9. It is declared as "negative" (i.e. normal) if it has a negative angiography or a D&F probability smaller than 0.1. The selection based on D&F stratification applies only to profiles without history of CAD. Out of our 1430 MPS cases, it remains 363 positive and 180 negative cases. Clinical interpretation is an alternative to "Class2". A case is classified as "positive" if it is considered as "ischemic", "fixed" or "mixed" by the nuclear physician and as "negative" if it is considered as "normal". In this way, we got 747 positive and 662 negative cases. Our severity criterion is the sum of all positive cases weighted by the correlation coefficient, normalized between a minimal correlation value and 1, minus the weighted negative cases. The minimal correlation level shall not be too high otherwise too few cases are defined as similar to perform statistical analysis. We have determined empirically that stable severity criteria can be obtained for a minimal correlation of 0.6 (see Figure 4).

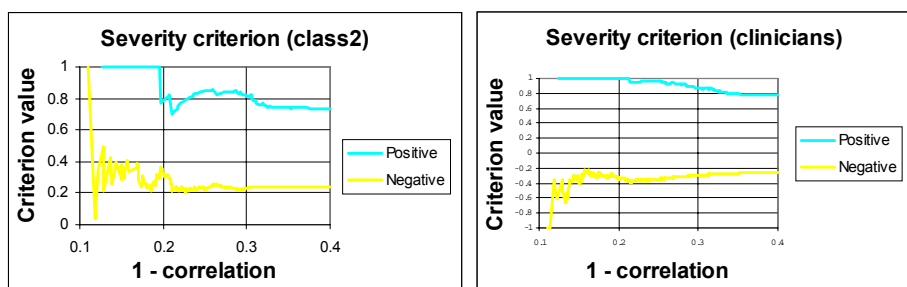


Figure 4: determining a minimal correlation level to define which cases to use in severity criteria.

6 Measuring ROC Curves of Severity Criteria Based on Similarity

To measure the operational characteristics of our severity criteria, we are measuring results only for cases with an established ground truth, that is, 131 angio-verified cases. Of course, a given case is not used in the computation of its own severity criteria to avoid bias. The same data set of 131 angio-verified cases was also blindly classi-

fied by 3 experts in a consensus session (Dr. G. DePuey, St. Lukes Hospital, New York, Dr. J. Vansant, Emory University, Atlanta, Dr. H. Yasuda, Mass General, Boston). We also compare the results with the performance of the clinicians that were studying these cases on a regular basis as well as with CardioMatch® software results.

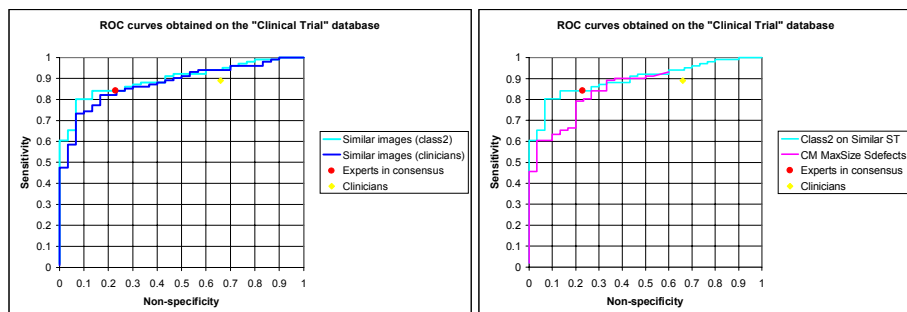


Figure 5: Left: comparing ROC for severity based on Class2 and severity based on Clinical diagnosis. Right: ROC curves to compare results obtained by clinicians, by experts in consensus, by CardioMatch® and by severity based on most similar patients (Class2).

We have first compared our two severity criteria to conclude that there is a small advantage in using Class2 classification instead of clinical diagnosis classification (see Figure 5). We have then compared results obtained by clinicians, by experts in consensus, by CardioMatch® and by severity based on most similar patients (Class2). For the same sensitivity than the clinicians (89%), both CardioMatch® and HeartPerfect are showing a much larger specificity (53% for CardioMatch®, 57% for HeartPerfect against 34% for clinicians). When compared to Experts in consensus (sensitivity = 84%, specificity = 77%), CardioMatch® is almost equivalent to experts (sensitivity = 84%, specificity = 73%, i.e. one case difference with respect to experts). For 84% sensitivity, HeartPerfect is reaching 87% specificity, which means a gain of 3 cases with respect to experts. We have also computed the distribution of negative and positive cases with respect to severity criteria (see Figure 6). It shows the discriminating power of the tests and also a slight advantage of the “class2” criterion over the criterion based on clinical information.

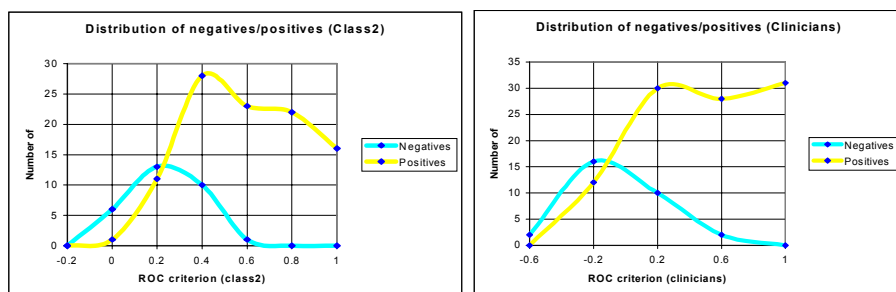


Figure 6: distribution of negative and positive cases with respect to severity criteria. Class2 (left) and clinical information (right).

7 Conclusion and Perspectives

We have described in this paper HeartPerfect, a prototype for data mining based on image similarity measures for the exploration of large databases of MPS studies. It relies on the geometric 3D normalization of all images performed in a pre-processing step. From a given patient case, it can output cases in the database presenting either the same clinical parameters, similar images based on image correlation measurements, or patients with similar perfusion defects. By computing statistics on most similar patients, we demonstrate that such a system is able to achieve a classification into CAD diseased or normal with a better performance than physicians in clinical routine and slightly superior performance than experts in consensus. The interest of HeartPerfect is that it can evolve simply through the enrichment of the database; more cases would allow restricting the search of similar images to subclasses of patients presenting similar clinical parameters. Up to now, we have used our system for Computer Aided Diagnosis only. In the future, we plan to enrich the database with information about the different outcomes of the patients, which would allow deriving statistics on possible outcome and ultimately, could allow inferring prognosis information.

8 Reference

1. J. Declerck, J. Feldmar, M.L. Goris and F. Betting, "Automatic Registration and Alignment on a Template of Cardiac Stress & Rest SPECT Images", in *Mathematical Methods in Biomedical Image Analysis*, June 1996.
2. M.L. Goris, W. Pace, M. Petersen and A. Kwan, "The Effect of Lesion Modulation on the Operating Characteristics of Myocardial Perfusion Studies", EANM'98, Berlin, Germany, August 1998.
3. M.L. Goris, W.M. Pace, J-P. Thirion, B. Hotz and P. Similon, "Case specific reference images for patient and acquisition variables in Myocardial Perfusion Analysis", EANM'99, Barcelona, Spain, Vol. 26, Num. 9, September 1999.
4. A. Guimond, G. Subsol and J-P. Thirion, "Automatic MRI Database Exploration and Applications", in *Int. J. on Pattern Recognition and Artificial Intelligence (IJPRAI)*, Vol. 11, Num. 8, pages 1345—1365, December 1997.
5. G.A. Diamond and J.S. Forrester, "Analysis of Probability as an Aid in the Clinical Diagnosis of Coronary-Artery Disease", in the *New England Journal of Medicine*, Vol. 300, Num. 24, pages 1350—1358, June, 1979.
6. A. Venot, J.F. Lebruchec, J.L. Golmard and J.C. Roucayrol, "An Automated Method for the Normalization of Scintigraphic Images", in the *Journal of Nuclear Medicine*, Vol. 24, pages 529—531, 1983.
7. C. Nastar, M. Mitschke, and C. Meilhac. "Efficient query refinement for image retrieval". In *Computer Vision and Pattern Recognition (CVPR '98)*, Santa Barbara, June 1998.
8. S.Aksoy and R.Haralick, "Textural features for image database retrieval", in *Proceedings of CVPR, Workshop on Content-Based Access of Image and Video Libraries*, 1998.