

RESEARCH ARTICLE

Integrating Dimension Reduction and Out-of-Sample Extension in Automated Classification of Ex Vivo Human Patellar Cartilage on Phase Contrast X-Ray Computed Tomography

Mahesh B. Nagarajan^{1*}, Paola Coan^{2,3,4}, Markus B. Huber¹, Paul C. Diemoz^{3,4}, Axel Wismüller^{1,2}

1 Departments of Imaging Sciences and Biomedical Engineering, University of Rochester Medical Center, Rochester, New York, USA, **2** Faculty of Medicine and Institute of Clinical Radiology, Ludwig Maximilian University, Munich, Germany, **3** Faculty of Physics, Ludwig Maximilian University, Munich, Germany, **4** European Synchrotron Radiation Facility, Grenoble, France

* mahesh.nagarajan@rochester.edu



OPEN ACCESS

Citation: Nagarajan MB, Coan P, Huber MB, Diemoz PC, Wismüller A (2015) Integrating Dimension Reduction and Out-of-Sample Extension in Automated Classification of Ex Vivo Human Patellar Cartilage on Phase Contrast X-Ray Computed Tomography. PLoS ONE 10(2): e0117157. doi:10.1371/journal.pone.0117157

Academic Editor: Qinghui Zhang, University of Nebraska Medical Center, UNITED STATES

Received: June 16, 2014

Accepted: December 18, 2014

Published: February 24, 2015

Copyright: © 2015 Nagarajan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: This research was funded in part by: 1. the National Institutes of Health (NIH) Award R01-DA-034977 (MBN MBH AW); 2. the Clinical and Translational Science Award 5-28527 within the Upstate New York Translational Research Network (UNYTRN) of the Clinical and Translational Science Institute (CTSI), University of Rochester (MBN MBH AW); 3. the Center for Emerging and Innovative Sciences (CEIS), a NYSTAR-designated Center for

Abstract

Phase contrast X-ray computed tomography (PCI-CT) has been demonstrated as a novel imaging technique that can visualize human cartilage with high spatial resolution and soft tissue contrast. Different textural approaches have been previously investigated for characterizing chondrocyte organization on PCI-CT to enable classification of healthy and osteoarthritic cartilage. However, the large size of feature sets extracted in such studies motivates an investigation into algorithmic feature reduction for computing efficient feature representations without compromising their discriminatory power. For this purpose, geometrical feature sets derived from the scaling index method (SIM) were extracted from 1392 volumes of interest (VOI) annotated on PCI-CT images of ex vivo human patellar cartilage specimens. The extracted feature sets were subject to linear and non-linear dimension reduction techniques as well as feature selection based on evaluation of mutual information criteria. The reduced feature set was subsequently used in a machine learning task with support vector regression to classify VOIs as healthy or osteoarthritic; classification performance was evaluated using the area under the receiver-operating characteristic (ROC) curve (AUC). Our results show that the classification performance achieved by 9-D SIM-derived geometric feature sets (AUC: 0.96 ± 0.02) can be maintained with 2-D representations computed from both dimension reduction and feature selection (AUC values as high as 0.97 ± 0.02). Thus, such feature reduction techniques can offer a high degree of compaction to large feature sets extracted from PCI-CT images while maintaining their ability to characterize the underlying chondrocyte patterns.

Advanced Technology (MBN MBH AW); and 4. the DFG Cluster of Excellence Munich - Centre for Advanced Photonics (MAP), Munich, Germany (PC PCD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funding sources had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

Osteoarthritis (OA) is now established as one of the leading causes of disability worldwide [1–3]. This disease is characterized by loss of articular cartilage, thickening of the underlying subchondral bone, and osteophyte formation [4]. Given that monitoring OA progression for purposes of patient health evaluation and response-to-therapy assessment are currently of significant interest, it would be desirable to have an imaging modality that could provide early detection and visualization of any degenerative modifications to cartilage [5–10]. Several imaging techniques are currently under investigation for their ability to assess cartilage health, eg. delayed gadolinium-enhanced MR imaging of cartilage (dGEMRIC) [11], ^{23}Na MRI [12], $T1\rho$ [13], GAG chemical exchange saturation transfer (gagCEST) [14] etc. These techniques focus on quantifying cartilage matrix composition where changes in water and collagen content, and loss in glycosaminoglycan (GAG) content have been previously identified as early signs of cartilage degeneration [13].

In this context, phase contrast X-ray computed tomography (PCI-CT) has recently emerged as a novel imaging modality that can visualize the internal architecture of the cartilage matrix at micro-meter scale resolution. Rather than rely on bio-chemical markers, PCI exploits the phase contrast effect associated with X-ray refraction in soft tissue, which is more pronounced than conventional absorption contrast in cartilage, as previously shown in [15]. Analysis of PCI-CT images acquired from ex vivo patellar cartilage specimens highlighted differences in chondrocyte organization between healthy and osteoarthritic cartilage matrix. Specific differences were noted in the radial zone where healthy specimens exhibited chondrocyte alignment (known as Benninghoff's arch [16]) while osteoarthritic specimens presented disorganized chondrocyte clustering throughout the matrix [15]. The high spatial resolution afforded by PCI-CT enables the use of texture analysis methods based on statistics (gray-level co-occurrence matrices or GLCM), topology (Minkowski Functionals), geometry (Scaling Index Method), etc to characterize these differences, as pursued in previous studies [17, 18]. Such textural approaches provide quantitative measures that could potentially serve as imaging markers for detecting and quantifying OA-induced degenerative changes to the cartilage matrix.

Textural approaches involving topological or geometrical features, as outlined in previous work [17, 18], provide a detailed characterization of the cartilage matrix through extraction of large feature sets. However, the extraction of too many features poses problems as it can contribute to overall deterioration in classification performance of machine learning algorithms, referred to as the so-called curse of dimensionality [19]. It has also been suggested that irrelevant or noisy features can adversely affect classification performance. Such problems highlight the need for employing some form of feature reduction to obtain an efficient representation of the original feature space while simultaneously maintaining adequate separability between the two classes of patterns, i.e. healthy and osteoarthritic. Feature reduction has been previously achieved with feature selection algorithms in the context of lung [20], breast [21, 22] etc. In such approaches, the original feature set is reduced through explicit exclusion of features either redundant in information content or irrelevant to the classification task. More recently, dimension reduction has also been proposed as an alternative to feature selection in the context of breast lesion classification [23–25]. Dimension reduction allows for an algorithmic weighting of all features in the original set while computing a newer smaller feature set.

We are specifically focused on evaluating the classification performance of feature sets extracted from post-processed phase contrast X-ray CT data for purposes of establishing imaging markers that can quantify OA-affected cartilage tissue. To that end, this specific study aimed at analyzing the impact of incorporating feature reduction on the performance achieved with previously proposed geometrical feature sets [18] in classifying healthy and osteoarthritic cartilage

tissue on PCI-CT. For this purpose, we present a new CADx methodology in this work where dimension reduction is integrated in conjunction with out-of-sample extension. Dimension reduction is applied to the training subset of the feature sets extracted from patellar cartilage VOIs; corresponding low-dimension representations for the test set are computed using out-of-sample extension techniques. Thus, a strict separation between training and test sets is maintained in our methodology, which is crucial for the supervised learning step in such automated classification tasks. This separates our work from previous attempts described in [23, 24] where dimension reduction was applied to the entire dataset prior to machine learning, which violates this training-test separation requirement. Our improved CADx methodology is described in detail in the following sections.

Data

Patellar Cartilage Samples

Age of the donor, macroscopic visual inspection and probing of the cartilage surface at autopsy were taken into account for selection of patellae. Donors older than 40 years were a priori excluded for harvest of normal samples while no constraint in age was imposed on potential donors for osteoarthritic samples. A smooth, white, and shiny surface present across the entire patellar cartilaginous surface and prompt resilience to manually performed focal indentation probing were criteria that defined macroscopically normal cartilage. Lack of these criteria in addition to visually perceived defects in the joint surface were used to select osteoarthritic samples. Based on these inclusion criteria, 2 healthy and 3 osteoarthritic cylinder-shaped osteochondral samples (diameter: 7 mm) were extracted within 48 hours postmortem from the lateral facet of 4 human patellae using a shell auger. Cylinders were trimmed to a total height of 12 mm including the complete cartilage tissue and the subchondral bone. The samples were continuously rinsed by 0.9% saline during extraction, trimming and removal of soiling from sawing. During image acquisition, samples were dipped into a 10% formalin solution.

PCI Experimental Setup

The image acquisition used the analyzer-based imaging (ABI) PCI technique, which has been previously demonstrated as highly sensitive to small phase variations [26]. The setup consisted of a parallel quasi-monochromatic X-ray beam, used to irradiate the sample, and of a perfect crystal, the analyzer, placed between the sample and the detector [27]. The analyzer acted as an angular filter of the radiation transmitted through the object and only the X-rays traveling in a narrow angle range close to the Bragg condition were diffracted onto the detector. Before being detected, the beam was modulated by the angle-dependent reflectivity of the crystal (rocking curve), which had a full width at half maximum (FWHM) typically of the order of a few microradians. All images were acquired at the half maximum position on one slope of the rocking curve (50% position), which was chosen to achieve the best sensitivity. Further details of this ABI technique can be found in [15, 28].

Experiments were performed at the Biomedical Beamline (ID17) of the European Synchrotron Radiation Facility (ESRF, France). Quasi-monochromatic X-rays of 26 keV were selected from the highly collimated X-ray beam by means of a double Si (111) crystal system and an additional single Si (333) crystal [29]. The emerging refracted and scattered radiation from the sample was analyzed with a Si (333) analyzer crystal. The imaging detector used was the Fast Readout Low Noise (FRLoN) CCD camera developed at the ESRF [30]. The X-rays are converted to visible light by a 60 μm thick Gadox fluorescent screen; this scintillation light is then guided onto a 2048×2048 pixel 14×14 m^2 CCD (Atmel Corp, US) by a lens-based system. The effective pixel size at the object plane was 8×8 μm^2 .

Tomographic Image Reconstruction

In order to acquire tomographic images with our PCI experimental setup, the cartilage samples were rotated about an axis perpendicular to the incident laminar beam. At the end of each rotation, the sample was displaced along this axis to enable imaging of a different region. Unlike conventional CT imaging, the beam and detector were kept stationary. To reduce the effects of any spatial and temporal X-ray beam inhomogeneities, we then performed a flat field normalization for each angular projection image. A direct Hamming filter backprojection (FBP) algorithm was used for reconstructing tomographic images [31]. For data analysis, coronal slices were reconstructed from the original data and subject to edge-preserving median filtering with a [5 5 5] sliding window to smoothen noise artifacts. An example image acquired from one healthy and one osteoarthritic specimen is shown in Fig. 1.

Pattern Annotation

Chondrocyte patterns were annotated with 3D cubic volumes of interest (VOI) in the radial zone of the cartilage matrix on the reconstructed PCI-CT images of all five specimens. 842 VOIs were annotated in total, of which 455 were osteoarthritic and 387 were healthy. The annotations were made using a cube of 25×25×25 pixels; the choice of VOI size was determined empirically based on previous work [18].

Methods

Ethics Statement

The institutional review board (IRB) of the Ludwig Maximilian University, Munich, Germany waived the need for ethical approval for this study since it involved retrospective analysis of anonymized tissue samples and imaging data collected from donors postmortem.

Overview

Fig. 2 presents the CADx methodology proposed and evaluated in this study. Different components of this system are described in this section.

Feature extraction is achieved with the use of novel geometrical features derived from the Scaling Index Method (SIM). While originally developed for analyzing multi-dimensional

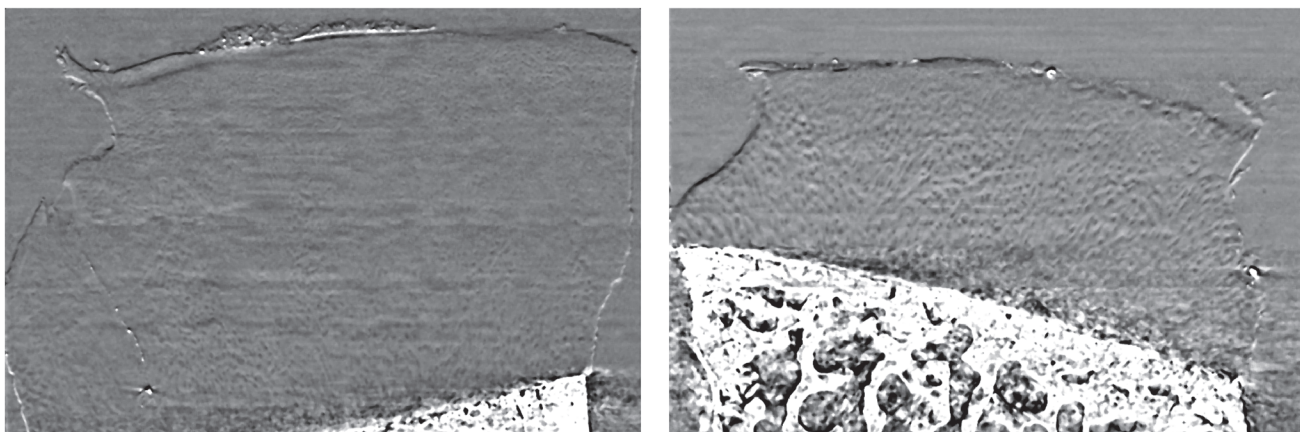


Fig 1. Coronal reconstruction created from axial PCI-CT images of a healthy (left) and osteoarthritic (right) cartilage specimens. Our interest is in the radial zone where chondrocyte organization is distinctly different in these two specimens.

doi:10.1371/journal.pone.0117157.g001

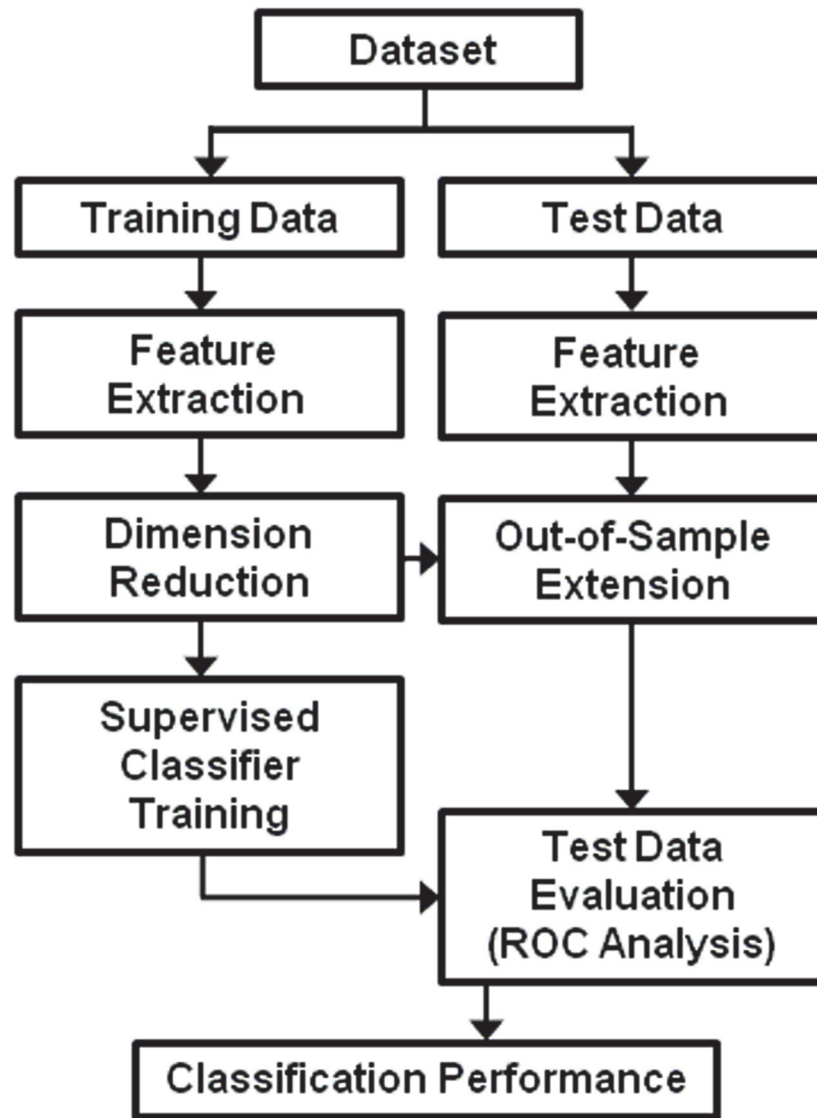


Fig 2. An overview of our CADx methodology proposed in this study. Our proposed methodology limits the application of dimension reduction to the training data alone, thus preserving the integrity (independence) of the test set. Low-dimension representations for the test set are obtained through out-of-sample extension.

doi:10.1371/journal.pone.0117157.g002

arbitrary point distributions through evaluation of the surrounding structural neighborhood [32], SIM has since been extended for estimating local scaling properties (or local dimension) of the gray-level intensity map within an annotated VOI [33]. In this work, texture analysis using SIM is pursued because of its suitability to the task of classifying between healthy and osteoarthritic cartilage, as previously demonstrated in [18].

The extracted feature vectors are then separated into training and test sets. The high-dimension feature vectors in the training set alone are subject to dimension reduction. While a wide variety of dimension reduction algorithms are described in the literature, we showcase our CADx methodology by focusing on a balanced selection of dimension reduction techniques (with respect to algorithmic properties) principal component analysis or PCA (non-parametric, linear) [34], Sammon’s mapping (classical gradient descent, non-linear) [35], t-distributed

stochastic neighbor embedding or t-SNE (global optimization, non-linear) [36] and exploratory observation machine or XOM (local optimization, non-linear) [37]. The corresponding low-dimension representations for the test set are obtained using out-of-sample extension techniques. In particular, we investigate the use of Shepard’s interpolation [38] and function approximation with a generalized radial basis function neural network (GRBF-FA) [39]. For comparison with dimension reduction, feature selection through evaluation of mutual information criteria [20] is also used.

Feature reduction is followed by supervised learning and classification, which is achieved through support vector regression (SVR) [40]. These processing steps were used to evaluate the classification performance achieved with our proposed CADx methodology of maintaining training-test data separation while applying dimension reduction. Individual components of this system are described in further detail in the following sub-sections.

Texture Analysis

In a specific VOI, all N voxels are represented by a 4-D vector $x_i = (x_i, y_i, z_i, g_i)$, $i = 1, 2, \dots, N$, consisting of three spatial dimensions (x_i, y_i, z_i) and voxel gray-level intensity $g_i = g(x_i, y_i, z_i)$. A unit scaling constant was used to define the relationship between the spatial and intensity dimensions of each voxel. The application of SIM for a given scale r can be treated as an image transformation where each voxel within the VOI is assigned a local scaling property $\alpha_i = \alpha(x_i, r)$. This scaling property reflects the structural and geometrical properties of the surface formed in the voxel neighborhood defined by r . We use a previously proposed estimator for α that uses a Euclidean distance metric and Gaussian shaping functions, i.e.,

$$\alpha(x_i, r) = \frac{2 \sum_{j=1}^N (d_{ij}/r)^2 e^{-(d_{ij}/r)^2}}{\sum_{j=1}^N e^{-(d_{ij}/r)^2}}, \tag{1}$$

where d_{ij} is the Euclidean distance between pixel x_i and neighboring pixel x_j and r is the radius of the Gaussian neighborhood [33]. After the SIM transformation is computed, the resulting distribution of α -values reveals non-linear structural information of the gray-level patterns annotated in the VOI. Nine quantiles (10th, 20th, . . . 90th) of this distribution were computed and used as a 9-D geometrical feature vector. The neighborhood radius was fixed as $r = 1$ based on previous work [18]. This is further illustrated in Fig. 3 using PCI-CT VOI examples.

Feature Reduction—Dimension Reduction

The goal of dimension reduction in this study was to obtain low-dimension representations of high-dimension feature vectors for subsequent classification. Specifically, we investigated the classification performance achieved with such representations of dimensions 2, 3, and 5, using the following methods.

Principal component analysis (PCA): PCA is an orthogonal linear transform that maps the original feature space to a new set of orthogonal coordinates or principal components [34]. This transform is defined in such a manner that the first principal component accounts for highest global variance, and subsequent principal components account for decreasing amounts of variance. The corresponding low-dimension representations of the SIM-derived geometric feature vectors can be determined by including the appropriate number of principal components (2, 3 or 5 in this study).

Sammon’s mapping: Sammon’s mapping establishes a point mapping relationship between high-dimension feature vectors and a low-dimension space so that inter-point distances in the

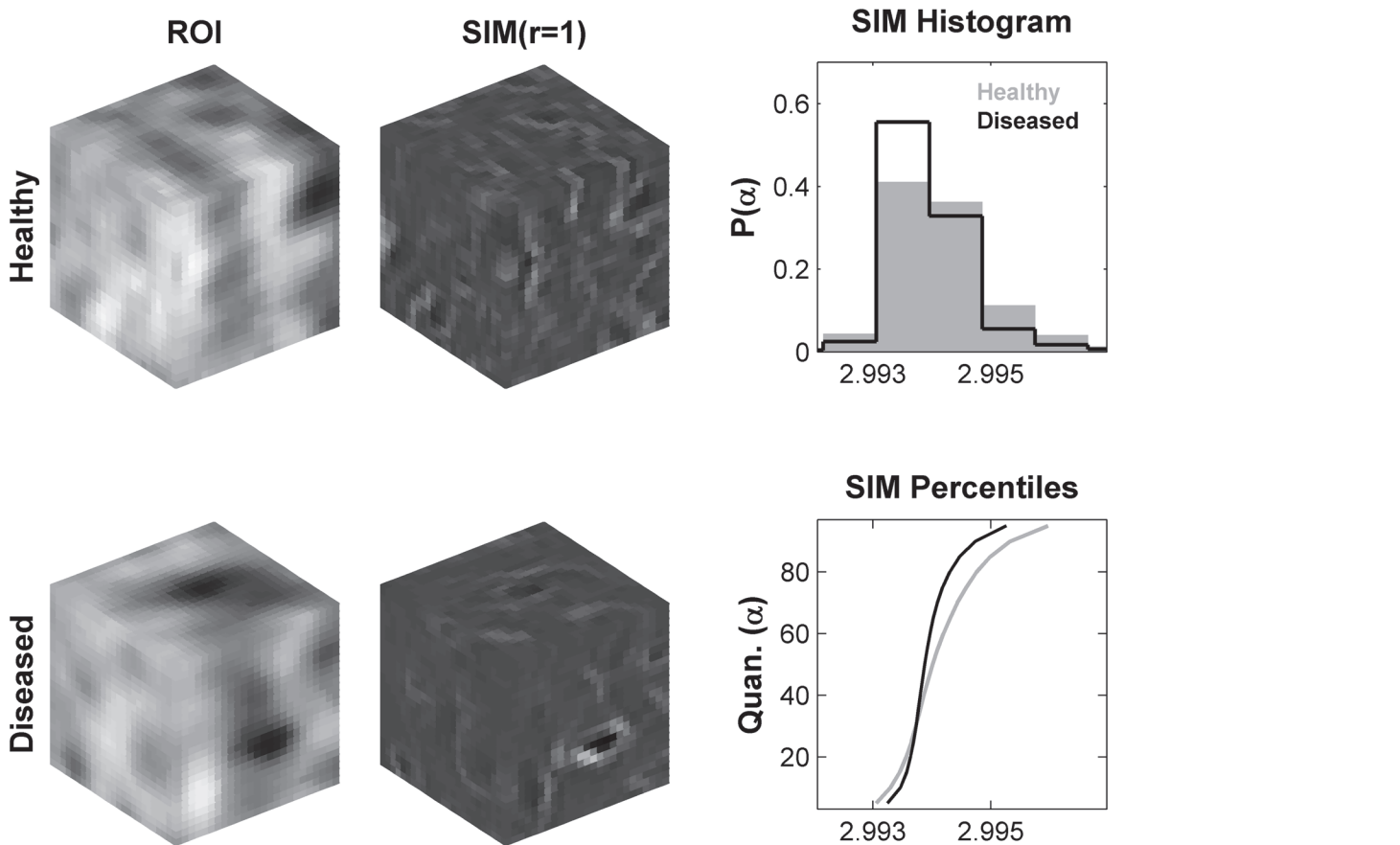


Fig 3. An illustration of the SIM feature extraction process. Examples of a normal and osteoarthritic VOIs (left), and their corresponding SIM transformations for radius $r = 1$ (middle). In the SIM transformations, dark regions correspond to lower magnitudes of α while brighter regions reflect higher magnitudes of α . The distribution of α -values from each SIM transformation are represented by histograms (top right) and by 9 percentiles (10^{th} – 90^{th}) (bottom right).

doi:10.1371/journal.pone.0117157.g003

high-dimension space approximate the corresponding inter-point distances in the low-dimension space [35].

Let $X_i, i = 1, 2, \dots, N$, represent a set of high-dimension feature vectors and $Y_i, i = 1, 2, \dots, N$, their corresponding low-dimension representations. The cost function E , which represents how well the low-dimension representations Y_i represent the feature vectors X_i , is given by—

$$E = \frac{1}{\sum_{i < j} D_{ij}} \sum_{i < j} \frac{(D_{ij} - d_{ij})^2}{D_{ij}}, \tag{2}$$

where the distance between any two points X_i and X_j is represented by D_{ij} , and the distance between any two points Y_i and Y_j , by d_{ij} . A steepest descent procedure is used for minimizing E . The implementation of this algorithm was taken from the self-organizing map (SOM) toolbox for MATLAB [41].

t-distributed stochastic neighbor embedding (*t*-SNE): Stochastic Neighbor Embedding (SNE) converts Euclidean distances between high-dimension texture feature vectors into conditional probabilities representing similarities; the closer the feature vectors, the higher the similarity

[36]. Once conditional probability distributions are established for both the high-dimension feature vectors and their corresponding low-dimension representations, the goal of the algorithm is to minimize the mismatch between the two.

Let $X_i, i = 1, 2, \dots, N$, represent a set of high-dimension feature vectors and $Y_i, i = 1, 2, \dots, N$, their corresponding low-dimension representations. Let $p_{j|i}$ be the condition probability that X_i selects X_j as a neighbor, assuming that neighbors were picked in proportion to their probability density under a Gaussian centered at X_i . Similarly, $q_{j|i}$ is the conditional probability in the low-dimension space. Minimizing the difference between $p_{j|i}$ and $q_{j|i}$ is achieved through minimization of the sum of Kullback-Leibler (KL) divergences over all feature vectors using a gradient descent method. The cost function is given by—

$$E = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \frac{p_{j|i}}{q_{j|i}}, \tag{3}$$

where P_i represents the conditional probability distribution over all other feature vectors given X_i , and Q_i represents the conditional probability distribution over all other low-dimension representations given Y_i .

t-SNE was developed as an improvement over SNE to further simplify cost function optimization and overcome the so-called crowding problem inherent to SNE [36]. Details pertaining to this algorithm and its cost function minimization can be found in [36], and a review of the algorithm can be found in [23, 37]. The t-SNE implementation used in this study was taken from the dimension reduction toolbox for MATLAB [42]. t-SNE has several free parameters, such as the degrees of freedom of the t-function, the number of iterations for which the cost function optimization is processed and perplexity, which can be defined as a smooth measure of the effective number of neighbors. All parameters were defined through default settings provided by the toolbox except for perplexity, which was optimized in the supervised learning step describe later.

Exploratory Observation Machine (XOM): As described in [43–45], XOM maps a finite number of data points X_i in a high-dimension space of dimension D to target points Y_i in the low-dimension embedding space of dimension d .

The initial setup of XOM involves—(1) defining the topology of the high-dimension data in the feature space through computation of distances $d(X_i, X_j)$ between feature vectors X_i , (2) defining a structure hypothesis represented by sampling vectors S_k in the low-dimension space, and (3) initializing output vectors Y_i , one for each input feature vector X_i . We use random samples from a uniform distribution for S_k in this study to enable occupation of the entire projection space. Once the initial setup was complete, the goal of the algorithm is to reconstruct the topology induced by the high-dimension feature vectors X_i through displacements of Y_i in the low-dimension space. Neighborhood couplings between feature vectors in the high-dimension space are represented by a cooperativity function h , which was modeled in this study as a Gaussian—

$$h(X_i, X'(S(t)), \sigma(t)) = e^{-\frac{(X_i - X'(S(t)))^2}{2\sigma^2(t)}}. \tag{4}$$

Here, $X'(S(t))$ represents the *best-match* for a input feature vector X_i . For a randomly selected sampling vector S , the *best-match* feature vector X' is identified by the criterion: $\|S - Y'\| = \min_i \|S - Y_i\|$. Once the *best-match* feature vector is identified, the output vectors Y_i are incrementally updated by a sequential adaptation step according to the learning rule

$$Y_i(t + 1) = Y_i(t) + \varepsilon(t)h(X_i, X'(S(t)), \sigma(t))(S(t) - Y_i(t)). \tag{5}$$

where t represents the iteration step, $\epsilon(t)$ is the learning rate and $\sigma(t)$ is a measure of neighborhood width taken into account by the cooperativity function h . In this study, both $\epsilon(t)$ and $\sigma(t)$ are changed in a systematic manner depending on the number of iterations by an exponential decay annealing scheme [45]. The algorithm is terminated when either the cost criterion is satisfied, or the maximum number of iterations is completed. The above sequential learning rule can be interpreted as a gradient descent step on a cost function for XOM, whose formal derivation can be found in [37]. The final position of Y_i represents the low-dimension representations of the high-dimension feature vectors.

We note three free parameters in this algorithm—(1) the learning parameter ϵ (2) the neighborhood parameter σ , and (3) the total number of iterations. As with t-SNE, default settings were specified for ϵ and number of iterations while σ was optimized in the supervised learning step describe later.

Out-of-Sample Extension

Since feature reduction through dimension reduction was restricted in its application to the training data alone, the test data were *out-of-sample* points. To obtain their corresponding low-dimension representations, the training set of high-dimension points X_i and their corresponding known low-dimension representations Y_i were used to define a mapping F such that $Y_i = F(X_i)$. This mapping F was then used to determine the low-dimension representations of the test set.

The goal of out-of-sample extension in this context was to create or approximate the mapping F . For a high-dimension feature vector X whose low-dimension representation is unknown, F can be treated as an interpolating function of the form

$$F(X) = \sum_{i=1}^N a_i(X) Y_i, \tag{6}$$

where a_i are the weights that define the interpolating function. We investigated two approaches to defining these weights.

Shepard’s Interpolation: This technique implements an inverse distance weighting approach in defining a_i described previously in [38], i.e.,

$$a_i(X) = \frac{1/d(X, X_i)^p}{\sum_{j=0}^N 1/d(X, X_j)^p}. \tag{7}$$

The power parameter p controls how points at different distances from X contributed to the computation of $F(X)$.

Generalized Radial Basis Function Neural Network Function Approximation (GRBF-FA): As an alternative to Shepard’s interpolation, the mapping F was approximated using a generalized radial basis function neural network. The weights a_i were defined as,

$$a_i(X) = \frac{e^{-\frac{(X-X_i)^2}{2\rho^2}}}{\sum_{j=0}^N e^{-\frac{(X-X_j)^2}{2\rho^2}}}, \tag{8}$$

which represented the activity of the hidden layer of the radial basis function network. The ρ parameter controlled the shape of the radial basis function kernel, and defined the neighborhood of feature vectors that contributed to the computation of $F(X)$.

Of the dimension reduction techniques investigated in this study, PCA was a special case that allowed for direct mapping of out-of-sample points into the low-dimension space and did not require any special out-of-sample extension.

We would also like to note here that such non-linear dimension reduction and out-of-sample extension techniques have free parameters which must be specified. While the typical approach is to optimize such parameters using different quality measures [35, 46, 47], we instead identified values for such free parameters that provided the best separation between the healthy and osteoarthritic classes of patterns, through cross-validation-based optimization conducted in the supervised learning step. We feel that our approach is justified since the best way to evaluate the quality of a lower-dimension projection is still unclear and under debate [46], and the end goal for dimension reduction in our study is classification and not visualization.

Feature Reduction—Feature Selection

Feature selection involves identifying a subset of features from the input feature space that makes the most relevant contribution to separating the two different classes of feature vectors in the supervised learning step. This study used mutual information analysis to identify a subset of features from the high-dimension feature vectors that best contributed to the pattern classification task.

Mutual information (MI) is a measure of general independence between random variables [19]. For two random variables X and Y , MI is defined as—

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \tag{9}$$

where entropy $H(\cdot)$ measures the uncertainty associated with a random variable. MI $I(X, Y)$ estimates how the uncertainty of X is reduced when Y has been observed. If X and Y are independent, their MI is zero.

For the dataset of ROIs used in this study, the MI between texture feature f_s , which is the feature stored in the s^{th} dimension of feature vector f , and the corresponding class labels y was calculated by approximating the probability density function of each variable using histograms $P(\cdot)$ —

$$I(f_s, y) = \sum_{n_c} \sum_{n_f} P(f_s, y) \log_2 \frac{P(f_s, y)}{P(f_s)P(y)}, \tag{10}$$

Here, the number of classes $n_c = 2$ was used; the number of histogram bins for the texture features n_f was determined adaptively according to

$$n_f = \log_2 N + 1 + \log_2(1 + \kappa \sqrt{N/6}), \tag{11}$$

where κ is the estimated kurtosis and N the number of ROIs in the data set [20].

Once the mutual information between each feature of the original feature set and the corresponding class labels was computed, those features with the highest mutual information were selected for subsequent classification. In this study, we investigated the classification performance achieved with 2, 3 and 5 features, as selected from the original feature set using mutual information criteria. To maintain training-test separation, the best features of the texture feature vectors were selected by evaluating the mutual information criteria of the training data alone.

Classification

The extraction of texture features and subsequent feature reduction was followed by a supervised learning step where the chondrocyte patterns were classified as healthy or osteoarthritic. In this work, support vector regression (SVR) with a linear kernel was used for the machine learning task [40]. The SVR implementation was taken from the libSVM library [48].

Owing to the practical limitations imposed by the small size of the patient population used in this study, we specified the following patient constraints to the supervised learning step—(1) ROIs from the same patient were not simultaneously used in both training and test sets, and (2) the same number of ROIs were used from every patient to ensure that the classifier did not get over-trained on patterns from a specific patient. Based on these constraints, each iteration of the supervised learning step involved randomly sub-sampling 200 ROIs from each of the five patients and randomly designating one each of the healthy and osteoarthritic subjects as the test set (the other samples comprised the training set). Such a strategy ensured that training sets used in different iterations of supervised learning were not identical despite patient constraints.

In the training phase, models were created from labeled data by employing a random sub-sampling cross-validation strategy where the training set was further split into 70% training samples and 30% validation samples. The purpose of the training phase was to determine the optimal parameters for the classifier, dimension reduction and out-of-sample extension algorithms that best captured the boundaries between the two classes of VOIs. The free parameters for the classifier used in this study were the cost parameter for SVR. Then, during the testing phase, the optimized classifier predicted the class of VOIs in the independent test set. A receiver-operating characteristic (ROC) curve was generated and used to compute the area under the ROC curve (AUC) which served as a measure of classifier performance on the independent test set. This process was repeated 50 times resulting in an AUC distribution for each feature set.

Statistical Analysis

A Wilcoxon signed-rank test was used to compare two AUC distributions corresponding to different texture features. Significance thresholds were adjusted for multiple comparisons using the Holm-Bonferroni correction to achieve an overall type I error rate (significance level) less than α (where $\alpha = 0.05$) [49, 50].

Texture, feature reduction, classifier and statistical analysis were implemented using Matlab 2010a (The MathWorks, Natick, MA).

Results

Evaluating Different Out-of-Sample Extension Methods

[Fig. 4](#) shows the classification performance achieved with the SIM-derived geometric feature vectors when processed with Sammon's mapping, XOM and t-SNE in conjunction with the two out-of-sample extension methods outlined earlier. No significant differences in performance are observed between Shepard's interpolation and GRBF-FA for both Sammon's mapping and XOM. However, with t-SNE, a significant improvement in performance was noted with Shepard's interpolation over GRBF-FA for 5-D projections of the original vectors ($p < 0.05$).

Comparing Dimension Reduction, Feature Selection and No Feature Reduction

[Table 1](#) shows the classification performance achieved with different feature reduction strategies pursued in this study. For each algorithm, the performance achieved with reduced feature

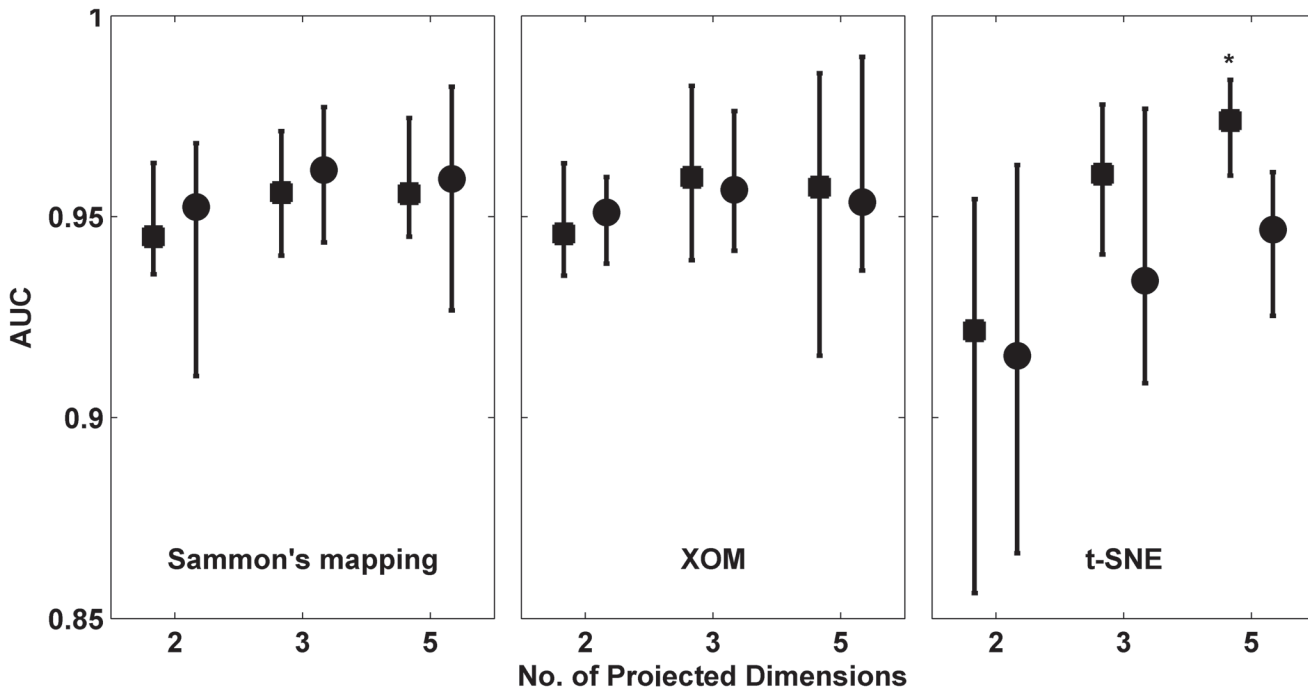


Fig 4. Comparison of classification performance in AUC (mean ± std) achieved with Sammon's mapping, XOM and t-SNE algorithms when used in conjunction with different out-of-sample extension techniques, i.e. Shepard's interpolation (square) and GRBF function approximation (circle). For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. Comparisons where the performance achieved with Shepard's interpolation were significantly better than those with GRBF-FA ($p < 0.05$) are marked with an asterisk.

doi:10.1371/journal.pone.0117157.g004

Table 1. Classification performance achieved with different feature reduction techniques.

Algorithm	Projected Dim.	AUC
PCA	2	0.97 ± 0.02
	3	0.97 ± 0.02
	5	0.97 ± 0.02
Sammon's Mapping	2	0.94 ± 0.03
	3	0.96 ± 0.02
	5	0.96 ± 0.02
XOM	2	0.95 ± 0.02
	3	0.96 ± 0.03
	5	0.94 ± 0.06
t-SNE	2	0.90 ± 0.09
	3	0.94 ± 0.08
	5	0.97 ± 0.02
Mutual Information	2	0.97 ± 0.01
	3	0.97 ± 0.02
	5	0.95 ± 0.02
none	9	0.96 ± 0.02

Classification performance in AUC (mean ± std) achieved with different dimension reduction and feature selection techniques investigated in this study. The last row shows the performance achieved when no feature reduction algorithm is applied, i.e. the original feature set is used for the classification task.

doi:10.1371/journal.pone.0117157.t001

sets of dimensions 2, 3 and 5 were evaluated and compared. For Sammon's mapping, XOM and t-SNE, Shepard's interpolation was used to obtain reduced feature representations of the independent test set.

The classification performance achieved with the original 9-D SIM-derived geometrical feature set, i.e. with no feature reduction strategy applied, was 0.96 ± 0.02 . When reducing this feature set to a 2-D representation, comparable classification was achieved by PCA (dimension reduction) and mutual information (feature selection). Other dimension reduction strategies such as Sammon's mapping, XOM and t-SNE were significantly outperformed ($p < 0.05$). However, for 3-D and 5-D representations, all dimension reduction and feature selection strategies yielded a comparable classification performance to the original feature set.

Discussion

Feature reduction strategies such as dimension reduction or feature selection have been previously proposed in computer-aided diagnosis (CADx) applications for obtaining efficient representations of large feature sets extracted from patterns of interest on medical images [20–25]. One advantage of identifying a reduced feature set representation of a large feature set is the reduction in processing time of the supervised learning step, as noted in this study (9-D: 6.03s, 5-D: 5.23s, 3-D: 4.88s, 2-D: 4.79s). However, the primary purpose of feature reduction in studies with finite datasets with limited patient cohort (or number of ROIs) is to prevent over-training, which stems from using too many features to describe too few ROIs, also known as the so-called curse of dimensionality [19]. This study evaluates the impact of incorporating such algorithms in the process of extracting feature sets that characterize chondrocyte organization in the radial zone of the cartilage matrix, as visualized on PCI-CT, for purposes of automated classification. The motivation for our work stems from previous demonstrations of PCI-CT's ability to visualize structural details of the human patellar cartilage matrix with high spatial resolution [15]. This makes cartilage imaging with PCI-CT a suitable target for soft tissue characterization with novel texture features [17, 18]. Given that such textural characterization usually yields a large feature set, it is important to obtain efficient representations of these extracted features through some feature reduction strategy.

In this study, we demonstrated a new CADx methodology for automated classification of healthy and osteoarthritic cartilage that incorporates dimension reduction into CADx while simultaneously maintaining a strict separation between training and test sets. This differentiates our study from previous attempts at using dimension reduction in CADx where such algorithms were applied to the entire dataset [23, 24]. Such implementation compromises the integrity of the independent test set, since feature vectors belonging to the training and test sets are free to interact and influence the computation of low-dimension representations. Our new methodology maintains the required training-test set separation by applying dimension reduction to the training data alone; corresponding low-dimension representations for the test set are obtained through out-of-sample extension techniques. Our methodology explored the integration of dimension reduction techniques such as PCA, Sammon's mapping, XOM and t-SNE in conjunction with out-of-sample extension techniques such as Shepard's interpolation and GRBF-FA for reducing the size of the originally extracted feature set.

Our results suggest that the high classification performance achieved with SIM-derived geometrical features (0.96 ± 0.02) can be maintained while substantially reducing the size of the original feature set. The fact that no statistically significant deterioration in performance is observed with the original feature set suggests room for further inclusion of other features when necessary. We specifically note that both dimension reduction through PCA and feature selection with mutual information were able to yield 2-D representations of the 9-D SIM feature set

extracted from the patellar cartilage ROIs without compromising the classification performance achieved. Non-linear dimension reduction techniques such as Sammon's mapping, XOM and t-SNE exhibited a small but significant deterioration in performance for 2-D projections of the original feature set, but exhibited comparable performance for 3-D and 5-D projections. The high classification performance noted with both the original feature set as well as the reduced feature set representations obtained with different techniques is further illustrated by Fig. 5, where distinct clusters of the healthy and osteoarthritic features are observed, albeit with poor separation as indicated by the corresponding Dunn's separation index [51]. Note that differences between the two clusters are only emphasized in the machine learning step; such visualizations are for exploratory analysis of the feature space only.

The high degree of compaction achieved by such techniques, i.e. reducing 9-D feature set to 2-D or 3-D, suggests tremendous potential for future application in computational tools for radiologists. As an example, content-based image retrieval (CBIR) could retrieve prior cases with similar patterns based on an annotated pattern in the current study. Such CBIR tools could rely on matching feature sets extracted from the current study to those previously extracted from other studies and stored in some database. In such a scenario, it's not surprising that the computational efficiency (in terms of processing speed, memory usage etc) improves when the feature sets are smaller in size. As long as such feature reduction approaches are only used in such ancillary support tools for radiologists in a clinical setting and not interpreted directly for evaluation of clinical findings, we anticipate minimal impact on clinical work flow in terms of information loss.

While we observe no significant differences in performance when using dimension reduction or feature selection for reducing the size of the original feature set in this study, their advantages and disadvantages are worth highlighting. Feature selection explicitly excludes features and results in a loss of information. Such losses could be relatively minimized in dimension reduction strategies where all features in the original set contribute to the final low-dimension representations. This has been previously observed when attempting to compact large feature sets (eg. 100-D) into very small representations (2-D/3-D) [25]. However, feature selection allows for identification of features that were selected as part of the reduced set. Dimension reduction results in the creation of new features, and the contributions of the original features to the reduced feature set is not readily interpretable. This will likely serve as an important criterion to consider while deciding upon which feature reduction strategy to pursue for a specific problem.

One must also note an inherent concern in integrating dimension reduction in its current form into CADx despite the promising results reported in this study. CADx aims to best separate different classes of feature vectors while dimension reduction attempts to best represent high-dimension data in a low-dimension space through some optimization paradigm (preservation of distances, similarities, topologies etc). These are essentially two independent optimization tasks with goals that are not guaranteed to align. One may explore dimension reduction techniques that incorporate some form of class discrimination while computing the low-dimension representations of the high-dimension feature vectors. Supervised dimension reduction variants of learning vector quantization approaches such as the neighbor retrieval visualize (NeRV) algorithm [46], generalized matrix learning vector quantization (GMLVQ) [52], or limited rank matrix learning vector quantization (LiRAM LVQ) [53], would be better suited to integration with our CADx methodology proposed in this study.

Finally, we acknowledge some limitations with the current study. To facilitate comparisons between different feature reduction algorithms, we arbitrarily fixed the sizes of the reduced feature sets to 2, 3 and 5. One could instead optimize for the smallest number of features that either yield the best classification performance or maintain the performance of the original

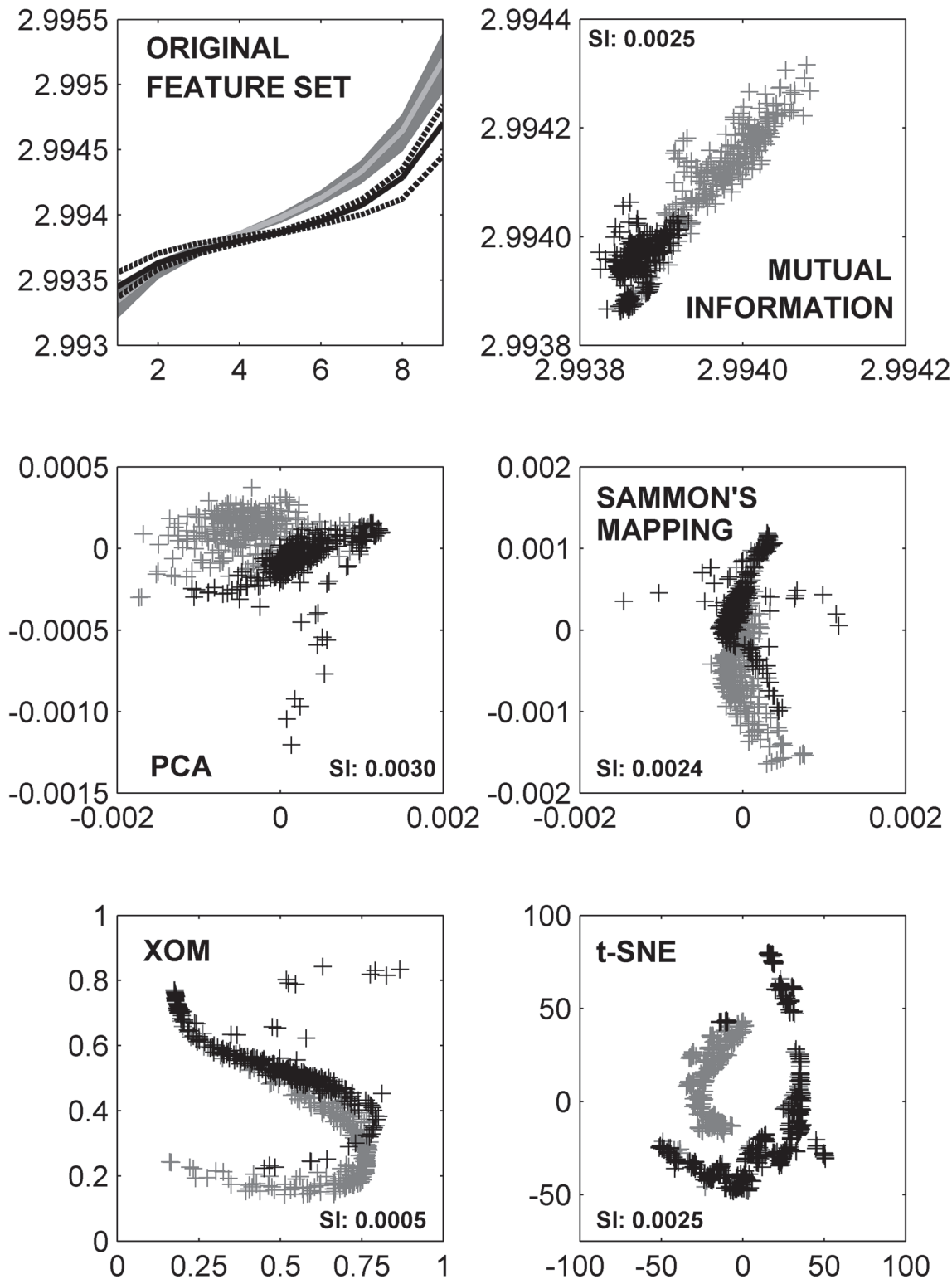


Fig 5. An exploration of the SIM feature space. In all plots, representations of feature vectors extracted from healthy VOIs are colored gray, those from osteoarthritic VOIs are colored black. Cluster separation is quantified with Dunn's separation index (SI), and specified in each plot. (Top Left) SIM feature vectors extracted from normal diseased ROIs area. The distribution of curves corresponding to each class is enclosed by the 25th and 75th percentile curves, the solid line represents the median curve. (Top Right) Plotting the 2-D reduced feature representation of the SIM feature set as obtained through evaluation of mutual information criteria. (Middle) Plots of 2-D projections of the SIM feature set obtained with PCA (left) and Sammon's mapping (right). (Bottom) Plots

of 2-D projections of the SIM feature set obtained with XOM (left) and t-SNE (right). As seen, here all feature reduction techniques yield discernible clusters of healthy and osteoarthritic VOIs, but with varying degrees of overlap.

doi:10.1371/journal.pone.0117157.g005

feature set. A small number of patients served as donors of the cartilage specimens for PCI-CT imaging and as a result, the classifier could be over-trained to the limited variations of healthy and osteoarthritic patterns found in these subjects. Future studies should include a larger patient cohort to ensure that the classifier is trained with a potentially larger variation of healthy and osteoarthritic patterns.

Conclusion

We demonstrate a CADx methodology with integrated feature reduction using either dimension reduction or feature selection in the research context of classifying healthy and osteoarthritic patellar cartilage annotated on PCI-CT images. We specifically outline a method to integrate dimension reduction in CADx while concurrently maintaining a strict training-test set separation required for supervised learning components. Our results suggest that both feature selection and dimension reduction could maintain the performance of the original pattern characterizing feature set while achieving a high degree of feature compaction. We hypothesize that such an approach would have significant practical advantages in a clinical setting as low-dimension representations of large feature sets extracted from annotated patterns can contribute to improved efficiency in terms of storage, processing speed etc. However, larger controlled trials need to be conducted in order to further validate the clinical applicability of our method.

Supporting Information

S1 Dataset. High-dimensional SIM-derived geometrical feature vectors and corresponding label data for the PCI-CT VOIs used in this manuscript.

(MAT)

Acknowledgments

This work was conducted as a practice quality improvement (PQI) project related to American Board of Radiology (ABR) maintenance of certificate (MOC) for Prof. Dr. Axel Wismüller. The authors would like to thank the ESRF for providing the experimental facilities and the ESRF ID17 team for assistance in operating the facilities. The following individuals are also acknowledged for their assistance with this work: Dr. Christian Glaser for his intellectual support, Dr. Emmanuel Brun for his assistance with the data sharing process, Benjamin Mintz for his assistance in developing the annotation tool used in this study, Dr. Annie Horng for her clinical insights and assistance with preparing this manuscript, and Prof. Dr. Maximilian Reiser, FACR, FRCR of the Department of Radiology, Ludwig Maximilian University, for his support. Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health (NIH).

Author Contributions

Conceived and designed the experiments: MBN AW. Performed the experiments: MBN. Analyzed the data: MBN. Contributed reagents/materials/analysis tools: PC MBH PCD AW. Wrote the paper: MBN AW. PCI-CT imaging: PC PCD.

References

1. Woolf A, Pfleger B (2003) Burden of major musculoskeletal conditions. *Bulletin of the World Health Organization* 81: 646–656. PMID: [14710506](#)
2. Yelin E (2003) Cost of musculoskeletal diseases: impact of work disability and functional decline. *Journal of Rheumatology* 68: 8–11. PMID: [14712615](#)
3. Maclean C, Knight K, Paulus H, Brook R, Shekelle P (1998) Costs attributable to osteoarthritis. *Journal of Rheumatology* 25: 2213–2218. PMID: [9818666](#)
4. Goldring M, Goldring S (2007) Osteoarthritis. *Journal of Cellular Physiology* 213: 626–634. doi: [10.1002/jcp.21258](#) PMID: [17786965](#)
5. Eckstein F, Wirth W, Nevitt M (2012) Recent advances in osteoarthritis imaging—the osteoarthritis initiative. *Nature Reviews: Rheumatology* 8: 622–630. PMID: [22782003](#)
6. Raya J, Hornig A, Dietrich O, Krasnokutsky S, Beltran L, et al. (2012) Articular cartilage: in vivo diffusion-tensor imaging. *Radiology* 262: 550–559. doi: [10.1148/radiol.11110821](#) PMID: [22106350](#)
7. Crema M, Roemer F, Marra M, Burstein D, Gold G, et al. (2011) Articular cartilage in the knee: current MR imaging techniques and applications in clinical practice and research. *Radiographics* 31: 37–61. doi: [10.1148/rg.311105084](#) PMID: [21257932](#)
8. Hunter D, Le Graverand MP, Eckstein F (2009) Radiologic markers of osteoarthritis progression. *Current Opinion in Rheumatology* 21: 110–117. doi: [10.1097/BOR.0b013e3283235add](#) PMID: [19339920](#)
9. Eckstein F, Glaser C (2004) Measuring cartilage morphology with quantitative magnetic resonance imaging. *Seminars in Musculoskeletal Radiology* 8: 329–353. doi: [10.1055/s-2004-861579](#) PMID: [15643573](#)
10. Glaser C, Faber S, Eckstein F, Fischer H, Springer V, et al. (2001) Optimization and validation of a rapid high-resolution T1-w 3D FLASH water excitation MRI sequence for the quantitative assessment of articular cartilage volume and thickness. *Magnetic Resonance Imaging* 19: 177–185. doi: [10.1016/S0730-725X\(01\)00292-2](#) PMID: [11358655](#)
11. Bashir A, Gray M, Boutin R, Burstein D (1997) Glycosaminoglycan in articular cartilage: in vivo assessment with delayed gd(DTPA)(2-)-enhanced MR imaging. *Radiology* 205: 551–558. doi: [10.1148/radiology.205.2.9356644](#) PMID: [9356644](#)
12. Reddy R, Li S, Noyszewski E, Kneeland J, Leigh J (1997) In vivo sodium multiple quantum spectroscopy of human articular cartilage. *Magnetic Resonance in Medicine* 38: 207–214. doi: [10.1002/mrm.1910380208](#) PMID: [9256099](#)
13. Stahl R, Luke A, Li X, Carballido-Gamio J, Ma C, et al. (2009) T1rho, T2 and focal knee cartilage abnormalities in physically active and sedentary healthy subjects versus early OA patients: a 3.0-Tesla MRI study. *European Radiology* 19: 132–143. doi: [10.1007/s00330-008-1107-6](#) PMID: [18709373](#)
14. Schmitt B, Zbyn S, Stelzeneder D, Jellus V, Paul D, et al. (2011) Cartilage quality assessment by using glycosaminoglycan chemical exchange saturation transfer and ²³Na MR imaging at 7 T. *Radiology* 260: 257–264. doi: [10.1148/radiol.11101841](#) PMID: [21460030](#)
15. Coan P, Bamberg F, Diemoz P, Bravin A, Timpert K, et al. (2010) Characterization of osteoarthritic and normal human patella cartilage by computed tomography X-ray phase-contrast imaging: A feasibility study. *Investigative Radiology* 45: 437–444. PMID: [20479648](#)
16. Benninghoff A (1925) Form und bau der gelenkknorpel in ihren beziehungen zur function. ii. der aufbau des gelenkknorpels in seinen beziehungen zur function. *Cell and Tissue Research* 2: 783–862.
17. Nagarajan M, Coan P, Huber M, Diemoz P, Glaser C, et al. (2013) Computer-aided diagnosis in phase contrast x-ray computed tomography for quantitative characterization of ex vivo human patellar cartilage. *IEEE Transactions on Biomedical Engineering* 60: 2896–2903. doi: [10.1109/TBME.2013.2266325](#) PMID: [23744660](#)
18. Nagarajan M, Coan P, Huber M, Diemoz P, Glaser C, et al. (2014) Computer-aided diagnosis for phase contrast x-ray computed tomography: Quantitative characterization of human patella cartilage with high-dimensional geometric features. *Journal of Digital Imaging* 27: 98–107. doi: [10.1007/s10278-013-9634-3](#) PMID: [24043594](#)
19. Duda R, Hart P, Stork D (2000) *Pattern Classification*. New York: Wiley-Interscience Publication.
20. Tourassi G, Frederick E, Markey M, Floyd C Jr (2001) Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics* 28: 2394–2402. doi: [10.1118/1.1418724](#) PMID: [11797941](#)
21. Nagarajan M, Huber M, Schlossbauer T, Leinsinger G, Krol A, et al. (2013) Classification of small lesions on breast MRI: Evaluating the role of dynamically extracted texture features through feature selection. *Journal of Medical and Biological Engineering* 33: 59–68. doi: [10.5405/jmbe.1183](#) PMID: [24223533](#)

22. Nagarajan M, Huber M, Schlossbauer T, Leinsinger G, Krol A, et al. (2013) Classification of small lesions in dynamic breast MRI: eliminating the need for precise lesion segmentation through spatiotemporal analysis of contrast enhancement. *Machine Vision and Applications* 24: 1371–1381. doi: [10.1007/s00138-012-0456-y](https://doi.org/10.1007/s00138-012-0456-y) PMID: [24244074](https://pubmed.ncbi.nlm.nih.gov/24244074/)
23. Jamieson A, Giger M, Drukker K, Li H, Yuan Y, et al. (2010) Exploring nonlinear feature space dimension reduction and data representation in breast CADx with laplacian eigenmaps and t-SNE. *Medical Physics* 37: 339–351. doi: [10.1118/1.3267037](https://doi.org/10.1118/1.3267037) PMID: [20175497](https://pubmed.ncbi.nlm.nih.gov/20175497/)
24. Jamieson A, Giger M, Drukker K, Pesce L (2010) Enhancement of breast CADx with unlabeled data. *Medical Physics* 37: 4155–4172. doi: [10.1118/1.3455704](https://doi.org/10.1118/1.3455704) PMID: [20879576](https://pubmed.ncbi.nlm.nih.gov/20879576/)
25. Nagarajan M, Huber M, Schlossbauer T, Leinsinger G, Krol A, et al. (2014) Classification of small lesions on dynamic breast MRI: Integrating dimension reduction and out-of-sample extension into CADx methodology. *Artificial Intelligence in Medicine* 60: 65–77. doi: [10.1016/j.artmed.2013.11.003](https://doi.org/10.1016/j.artmed.2013.11.003) PMID: [24355697](https://pubmed.ncbi.nlm.nih.gov/24355697/)
26. Diemoz P, Bravin A, Langer M, Coan P (2012) Analytical and experimental determination of signal-to-noise ratio and figure of merit in three phase-contrast imaging techniques. *Optics Express* 20: 27670–27690. doi: [10.1364/OE.20.027670](https://doi.org/10.1364/OE.20.027670) PMID: [23262715](https://pubmed.ncbi.nlm.nih.gov/23262715/)
27. Bravin A (2003) Exploiting the X-ray refraction contrast with an analyser: the state of the art. *Journal of Physics D: Applied Physics* 36: 24–29. doi: [10.1088/0022-3727/36/10A/306](https://doi.org/10.1088/0022-3727/36/10A/306)
28. Bravin A, Coan P, Suortti P (2013) X-ray phase-contrast imaging: from pre-clinical applications towards clinics. *Physics in Biology and Medicine* 58: R1–35. doi: [10.1088/0031-9155/58/1/R1](https://doi.org/10.1088/0031-9155/58/1/R1) PMID: [23220766](https://pubmed.ncbi.nlm.nih.gov/23220766/)
29. Fiedler S, Bravin A, Keyriläinen J, Fernández M, Suortti P, et al. (2004) Imaging lobular breast carcinoma: comparison of synchrotron radiation DEI-CT technique with clinical CT. *Physics in Medicine and Biology* 49: 175–188. doi: [10.1088/0031-9155/49/2/001](https://doi.org/10.1088/0031-9155/49/2/001) PMID: [15083665](https://pubmed.ncbi.nlm.nih.gov/15083665/)
30. Coan P, Peterzola A, Fiedler S, Ponchut C, Labiche J, et al. (2006) Evaluation of imaging performance of a taper optics CCD ‘FReLoN’ camera designed for medical imaging. *Journal of Synchrotron Radiation* 13: 260–270. doi: [10.1107/S0909049506008983](https://doi.org/10.1107/S0909049506008983) PMID: [16645252](https://pubmed.ncbi.nlm.nih.gov/16645252/)
31. Dilmanian F, Zhong Z, Ren B, Wu X, Chapman L, et al. (2000) Computed tomography of X-ray index of refraction using the diffraction enhanced imaging method. *Physics in Medicine and Biology* 45: 933–946. doi: [10.1088/0031-9155/45/4/309](https://doi.org/10.1088/0031-9155/45/4/309) PMID: [10795982](https://pubmed.ncbi.nlm.nih.gov/10795982/)
32. Jamitzky F, Stark W, Bunk W, Thalhammer S, Raeth C, et al. (2000) Scaling-index method as an image processing tool in scanning-probe microscopy. *Ultramicroscopy* 86: 241–246. doi: [10.1016/S0304-3991\(00\)00111-X](https://doi.org/10.1016/S0304-3991(00)00111-X) PMID: [11215629](https://pubmed.ncbi.nlm.nih.gov/11215629/)
33. Raeth C, Bunk W, Huber M, Morfill G, Retzlaff J, et al. (2002) Analysing large scale structure: I. weighted scaling indices and constrained randomization. *Monthly Notice of the Royal Astronomical Society* 337: 413–421. doi: [10.1046/j.1365-8711.2002.05829.x](https://doi.org/10.1046/j.1365-8711.2002.05829.x)
34. Hotelling H (1933) Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 24: 498–520. doi: [10.1037/h0071325](https://doi.org/10.1037/h0071325)
35. Sammon J (1969) A nonlinear mapping for data structure analysis. *IEEE Transactions on Computers* C-18: 401–409. doi: [10.1109/T-C.1969.222678](https://doi.org/10.1109/T-C.1969.222678)
36. van der Maaten L, Hinton G (2008) Visualizing data using t-SNE. *Journal of Machine Learning Research* 9: 2579–2605.
37. Bunte K, Hammer B, Villman T, Biehl M, Wismüller A (2011) Neighbor embedding XOM for dimension reduction and visualization. *Neurocomputing* 74: 1340–1350. doi: [10.1016/j.neucom.2010.11.027](https://doi.org/10.1016/j.neucom.2010.11.027)
38. Shepard D (1968) A two-dimensional interpolation function for irregularly-spaced data. In: Blue R, Rosenberg A, editors, *Proceedings of the 1968 23rd ACM National Conference*. New York: ACM, pp. 517–524.
39. Moody J, Darken C (2009) Fast learning in networks of locally-tuned processing units. *Neural Computation* 1: 281–294. doi: [10.1162/neco.1989.1.2.281](https://doi.org/10.1162/neco.1989.1.2.281)
40. Drucker H, Burges C, Kaufman L, Smola A, Vapnik V (1996) Support vector regression machines. *Advances in Neural Information Processing Systems* 9: 155–161.
41. Vesanto J, Himberg J, Alhoniemi E, Parhankangas J (2000) Self-organizing map in matlab: the som toolbox. In: *Proceedings of the Matlab DSP Conference*. pp. 35–40. Software available at <http://www.cis.hut.fi/projects/somtoolbox/>.
42. van der Maaten L, Postma E, van den Herik H (2009) Dimensionality reduction: A comparative review. *Tilburg University Technical Report TiCC-TR 2009-005*: 1–22.
43. Wismüller A (2009) A computational framework for nonlinear dimensionality reduction and clustering. In: Principe J, Miikkulainen R, editors, *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, volume 5629, pp. 334–343.

44. Wismüller A (2009) The exploration machine—a novel method for data visualization. In: Principe J, Miikkulainen R, editors, *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, volume 5629, pp. 344–352.
45. Wismüller A (2009) The exploration machine—a novel method for analyzing high-dimensional data in computer-aided diagnosis. In: Karssemeijer N, Giger M, editors, *Proceedings of SPIE*. Bellingham: SPIE, volume 7260, pp. 0G1–0G7.
46. Venna J, Peltonen J, Nybo K, Aidos H, Kaski S (2010) Information retrieval perspective to nonlinear dimensionality reduction for data visualization. *Journal of Machine Learning Research* 11: 451–490.
47. Lee J, Verleysen M (2009) Quality assessment of dimensionality reduction: Rank-based criteria. *Neurocomputing* 72: 1431–1443. doi: [10.1016/j.neucom.2008.12.017](https://doi.org/10.1016/j.neucom.2008.12.017)
48. Chang CC, Lin CJ (2011) LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* 2: 27:1–27:27. doi: [10.1145/1961189.1961199](https://doi.org/10.1145/1961189.1961199)
49. Wright SP (1992) Adjusted P-values for simultaneous inference. *Biometrics* 48: 1005–1013. doi: [10.2307/2532694](https://doi.org/10.2307/2532694)
50. Holm S (1979) A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6: 65–70.
51. Dunn C (1973) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics* 3: 32–57. doi: [10.1080/01969727308546046](https://doi.org/10.1080/01969727308546046)
52. Schneider P, Biehl M, Hammer B (2009) Adaptive relevance matrices in learning vector quantization. *Neural Computation* 21: 3532–3561. doi: [10.1162/neco.2009.11-08-908](https://doi.org/10.1162/neco.2009.11-08-908) PMID: [19764875](https://pubmed.ncbi.nlm.nih.gov/19764875/)
53. Bunte K, Schneider P, Hammer B, Schleif F, Villman T, et al. (2012) Limited rank matrix learning, discriminative dimension reduction and visualization. *Neural Networks* 26: 159–173. doi: [10.1016/j.neunet.2011.10.001](https://doi.org/10.1016/j.neunet.2011.10.001) PMID: [22041220](https://pubmed.ncbi.nlm.nih.gov/22041220/)