HHS Public Access

Author manuscript

Comput Biol Med. Author manuscript; available in PMC 2019 April 01.

Published in final edited form as:

Comput Biol Med. 2018 April 01; 95: 24–33. doi:10.1016/j.compbiomed.2018.01.008.

Deep Transfer Learning for Characterizing Chondrocyte Patterns in Phase Contrast X-Ray Computed Tomography Images of the Human Patellar Cartilage

Anas Z. Abidin[a,b,*], **Botao Deng**[c], **Adora M. DSouza**[c], **Mahesh B. Nagarajan**[d], **Paola Coan**[e,f], and **Axel Wismüller**[a,b,c,f]

[a]Department of Biomedical Engineering, University of Rochester Medical Center, Rochester, New York, USA

[b]Department of Imaging Sciences, University of Rochester Medical Center, Rochester, New York, USA

[c]Department of Electrical Engineering, University of Rochester Medical Center, Rochester, New York, USA

[d]Department of Radiological Sciences, University of California Los Angeles, Los Angeles, U.S.A

[e]European Synchrotron Radiation Facility, Grenoble, France

[f]Faculty of Medicine and Institute of Clinical Radiology, Ludwig Maximilians University, Munich Germany

Abstract

Phase contrast X-ray computed tomography (PCI-CT) has been demonstrated to be effective for visualization of the human cartilage matrix at micrometer resolution, thereby capturing osteoarthritis induced changes to chondrocyte organization. This study aims to systematically assess the efficacy of deep transfer learning methods for classifying between healthy and diseased tissue patterns. We extracted features from two different convolutional neural network architectures, CaffeNet and Inception-v3 for characterizing such patterns. These features were quantitatively evaluated in a classification task measured by the area (AUC) under the Receiver Operating Characteristic (ROC) curve as well as qualitative visualization through a dimension reduction approach t-Distributed Stochastic Neighbor Embedding (t-SNE). The best classification performance, for CaffeNet, was observed when using features from the last convolutional layer and the last fully connected layer (AUCs > 0.91). Meanwhile, off-the-shelf features from Inception-v3 produced similar classification performance (AUC > 0.95). Visualization of features from these layers further confirmed adequate characterization of chondrocyte patterns for reliably distinguishing between healthy and osteoarthritic tissue classes. Such techniques, can be

[*]Corresponding author. anas.abidin@rochester.edu (Anas Z. Abidin).

Disclosures

The authors do not have any financial interests to disclose.

potentially used for detecting the presence of osteoarthritis related changes in the human patellar cartilage.

Keywords

phase contrast imaging; patellar cartilage; deep transfer learning; convolutional neural network

1. Introduction

Phase-contrast X-ray computed tomography (PCI-CT) is an imaging technique capable of visualizing the internal architecture of tissues at micrometer resolution [1]. This acquisition methodology exploits the fact that X-rays are not just absorbed when passing through matter but also refracted,[2, 3] producing a more pronounced contrast when compared to conventional absorption based X-ray imaging modalities[4]. This allows PCI to be effective in imaging tissue types where the conventional absorption contrast is either unable to resolve the differences between soft tissue types, i.e., breast[5, 6], or is poor/absent, i.e., cartilage[1, 7, 8]. Although different imaging setups can be used, PCI with computed tomography, using the analyzer-based imaging (ABI) scheme [3, 9, 10] has been applied in different ex-*vivo* breast[5, 6], brain[11], and cartilage studies[1, 7]. There is a huge potential for such a technique to be used for early detection of degenerative cartilage structural changes associated with osteoarthritis (OA) [12], widely recognized as one of the leading causes of disability worldwide. The advanced techniques proposed for evaluation of OA, such as delayed gadolinium-enhanced MR imaging of cartilage (dGEM-RIC) [13], ^{23}Na MRI [14], T1ρ [15], GAG chemical exchange saturation transfer (gagCEST) [16] etc. do not possess the capability to visualize cartilage matrix structure at a cellular level. In this context, the adept visualization of the cartilage matrix using PCI-CT has been demonstrated previously [1]. It was observed that the systematic zonal architecture maintained in the cartilage matrix, as known from histology studies [17], was clearly visualized in samples obtained from healthy individuals. Specifically in the radial zone - defined based on established histopathological standards [18, 19] - the chondrocytes demonstrate an ordered arrangement (Fig. 1, top) which is understood to be progressively lost during disease progression and hence was not observed in the osteoarthritic samples. Instead, a more generalized clustering of cells throughout the matrix was observed (Fig. 1, bottom).

We have previously [20, 21] shown that PCI-CT images can be characterized effectively with 2D or 3D texture features, in a computer aided diagnostics framework. In this study, we explore the use of deep learning for characterizing chondrocyte organization of the cartilage matrix visualized in these images. Conventionally, features (such as those derived from Grey Level Co-occurance Matrices, GLCM [22]) extracted from images were used with a classification algorithm such as a support vector machine (SVM) to distinguish between classes of subjects. In these analyses, the feature are computed based on heuristic characteristics of the images and in general are performed independent of the classification algorithm. With the advent of deep learning techniques, focus has shifted from the use of traditional, so called “hand-crafted”, features to the use of networks which learn representations best suited for a task. There has been a tremendous growth in methods to go

along with vast improvements in performance of such systems in various computer vision tasks [23, 24, 25]. An increasing number of medical imaging studies [26] are using deep learning methods for recognition, classification, and segmentation tasks. As these neural-networks require large amounts of annotated data, which may often not be available for medical imaging, the use of *transfer learning* has been suggested. Here, networks pre-trained on large image databases can be adapted for the specific task at hand.

One approach is to treat Convolutional Neural Network (CNN) like a feature extractor *i.e.* using "off-the-shelf" feature representations from intermediate layers of CNN and use them with a commonly used classifier such as a SVM. Such features have shown remarkable performance at various visual recognition tasks [23] including image classification, attribute detection and image retrieval. In medical imaging, these approaches have been applied for the detection of a wide range of chest-related diseases [27] and pulmonary nodule in computed tomography scans [28]. A particular question that is not always answered in these studies is what level of representation is suited when using a convolutional neural network as a feature extractor. Although this issue has been addressed briefly [23, 29] in computer-vision literature, the conclusions may not be directly applicable to medical images, as their network was both trained and tested on natural image datasets. To address such a problem with our dataset, we aim to study the characterizations obtained using different internal layers of the network for distinguishing healthy and diseased tissue classes.

The other widely used technique for transfer learning, is to *fine-tune* a pre-trained network on a target medical imaging dataset[30]. The method is used to adapt the weights of the network based on using new training data. It is based on the accepted idea that in a deep learning network, the initial layers capture features that are generic (edges, orientations, simple text patterns etc.) and as we go deeper the features tend to get more abstract and tend to increase in specificity for the task they are being trained on. To explore this effect in detail, we fine-tune the weights of a pre-trained network and test its performance in a classification task.

This is one of the first studies exploring the application of deep transfer learning on phase contrast imaging data. We aim to investigate the effectiveness of CNNs in characterizing degenerative changes occurring due to osteoarthritis in PCI-CT images of the cartilage. We also try to address the following, technically relevant, questions in this work:

- How do off-the-shelf features from a CNN perform when classifying PCI-CT images acquired from healthy and osteoarthritic cartilage samples?
- In the commonly used architecture, CaffeNet (based on AlexNet); what is the adequate representation that can be extracted from the internal layers?
- Can fine-tuning with a small dataset help in improving the performance of tranfer learning methods?
- Futhermore, we explore the application of Inception-v3, an advanced CNN architecture which is not commonly used in medical imaging and compare its performance to CaffeNet.

- Can visual exploratory analyses shed further light on the descriptors obtained from the CNN?

2. Data

The imaging data used in this study is acquired from scans of five ex-*vivo* specimens of the cartilage extracted from the retropatellar joint which is understood to possess significant potential for enabling early detection of treatable osteoarthritic changes. Specimens were selected for scanning based on the following (inclusion) criteria: age of the donor (< 40 yrs of age), visual inspection of specimen (smooth, white and shiny surface present across the cartilage surface) and probing of the cartilage surface (prompt resilience to focal indentation). IRB was waived by the institutional review board of the Ludwig Maximilians University, Munich, Germany. The final sample consisted of five cylindrical osteochondral samples extracted within 48 hours postmortem from the patella. Based on histological standards [31], one diseased sample was graded as mildly osteoarthritic (grade 3) and two were assessed to have advanced OA (grade 4), while two samples did not show any osteoarthritic changes.

Details of the imaging setup have been described in detail in previous studies [1, 7]. We briefly summarize the process here. Phase contrast imaging was performed using an analyzer based imaging (ABI) setup which consists of an analyzer (a perfect crystal) placed between a parallel monochromatic X-ray beam source and the detector [32]. The analyzer acts as an angular filter of the radiation transmitted through the object, which can be modulated by the angle-dependent reflectivity of the crystal. A Si (333) crystal was used an analyzer. Highly collimated and Quasimonochromatic X-rays at 26 keV, generated using a 21-pole wiggler were used to irradiate the samples. The detector used was the Fast Readout Low Noise (FReLoN) CCD camera developed in-house [33]. Within this imaging setup, the source as well as the detector were kept stationary, with the sample being rotated to acquired an angular projection dataset over 360°, in 1° increments. An image volume of dimensions 1120 × 1124 × 805, with an effective pixel size of $8 \times 8 \mu m^2$ in the imaging plane, was obtained for each specimen and subsequently trimmed to eliminate background regions. Imaging was performed at the Biomedical Beamline (ID17) of the European Synchrotron Radiation Facility (ESRF, France).

It was reported in [1] that the osetoarthritic samples showed changes in chondrocyte organization in PCI-CT images, particularly in the radial zone. Hence, a total of 842 square ROIs (size 101 × 101 pixels, 439 osteoarthritic and 403 healthy), capturing chondrocyte patterns in the radial zone of the cartilage matrix, were annotated on the acquired PCI-CT images for all specimens. The ground truth was extracted using analysis performed by two independent observers using histopathology standards. To ensure a good sampling of the matrix and avoid subsequent over-fitting in the machine learning step, additional restrictions were imposed for ROI definition. Within a specimen, ROIs were defined atleast $32 \mu m$ apart, this ensured that same cluster of chondrocytes was not captured across multiple ROIs. Furthermore, to avoid ambiguous ROI definition and avoid class imbalance, ROIs in the OA samples were not defined in regions showing normal patterns of chondrocyte organization.

Detailed information regarding specimen preparation, imaging setup and ground truth extraction is avaliable in our previously published studies[1, 20].

3. Methods

In studies such as ours, where a large dataset is not available for training deep learning network from initialization, researchers have often used what is known as “transfer-learning” from pre-trained networks. Such models are made publicly available after training on a large annotated imaging database. For characterizing the patterns in the cartilage matrix, in this study we have used and compared the performance of two different pre-trained convolutional neural networks (CNNs): CaffeNet[34] and Inception-v3[35], which were trained on the Imagenet database.

3.1. Features from CaffeNet

3.1.1. Off-the-shelf—CaffeNet, an adaptation of the winning architecture AlexNet of ILSVRC 2012 [36], is one of the most widely used networks used in transfer learning. Its simple architecture, consisting of five convolutional layers (*conv*) and 3 fully connected layers (*fc*), facilitates easier benchmarking of classification performance for different CNNs. The trained network, trained on about 1 million images of the ImageNet database, has been made available by Jia et al.[34]. Each ROI from our study was scaled with bilinear interpolation to match the size of input layer (227 × 227) and a training set mean was subtracted in line with standard practice in deep learning studies [34].

It has been proposed that features in the intermediate layers capture adequate representations for transfer-learning studies[23] as these layers are neither too specific to the dataset the network was originally trained on, nor too general to not contain any representative information from images. To test whether such a rule would apply to our PCI-CT images, we extracted features from all layers of the CaffeNet. Henceforth, features from the 5 convolutional and 3 fully connected layers of the CaffeNet (excluding the normalization layer or pooling layers), are denoted as $conv_n$ ($1 \leq n \leq 5$) or fc_n ($6 \leq n \leq 8$), respectively.

Features from the convolutional layers are multidimensional and have a relatively large size (e.g. $conv_2$: 27 × 27 × 256, $conv_4$: 13 × 13 × 348), thereby increasing the computational burden. We, therefore, applied Global Average Pooling (GAP) to pool the feature maps for converting them to linear feature vectors (e.g. $conv_1$:55×55×96 into a vector 1×1×96). Although other strategies to pool features from the convolutional layers can be considered [37, 38] GAP has been widely applied in the development of deeper architectures. It also enforces correspondance between feature maps [39], does not require additional free parameter optimization, and is robust to spatial transformations.

3.1.2. Fine-tuning—To study the effect of fine-tuning, we replaced the final fully connected layer (fc_8) in the original CaffeNet with a 2 ouput layer corresponding to the healthy and osteoarthritic classification. The weights were randomly initialized based on a Gaussian distribution. During fine-tuning, we ensured that the errors were back-propagated to all layers of the network. The learning rate for the last layer was fixed at 0.001 with 0.9 momentum as this layer had to be learnt from scratch. However the learning rates for the

remaining layers were set at 0.0005 as only fine adjustments from a pre-trained state would be needed. The network was trained for 1500 iterations with a batch size of 256. The multinomial logistic loss function reached a low plateau after about 1000 training iterations. We report results obtained using the two output layer of the CNN.

3.2. Features from Inception-v3

We also tested the performance of an Inception-v3 network, a contemporary architecture for characterizing our images. This network is a scaled-up version of GoogLeNet, and augments various heuristic improvisations over CNN architectures such as AlexNet, OverFeat or Decaf previously used in medical imaging studies. The pretrained Inception-v3 model was obtained from the open source deep learning framework, Lasagne Table 1 shows the architecture of Inception-v3 model, the first few layers are similar to traditional CNN, including 5 convolutional layers and 2 max pooling layers. These are followed by 11 inception modules and a Global pooling layer. A softmax layer is used to generalize a score for each category.

The core idea of the Inception module is to use filters of multiple sizes with different receptive field, this allows for capturing local variablity and hence improved characterization. The original naive inception module in GoogLeNet performed convolution operations of three sizes, 1×1, 3×3, 5×5, and then combined them together to form a single feature matrix which is fed into the subsequent layers. Such operations when performed for larger networks quickly increase the amount of computations which can be prohibitively expensive. Improvements in the implementation of convolutions as suggested by [39], significantly circumvents this problem. These modules also serve to enhance the representative power of networks by reducing the correlations in the activations of nearby neurons. Further details of the Inception module are discussed elsewhere [40]. Such improvements have allowed for significant enhancements in the representative power of neural networks as evidenced by the performance of Inception networks in ImageNet competition(s).

For each image, the training set mean image was subtracted, and input image is resized to match the input layer dimension of Inception-v3. We extracted features from the last 10 inception modules, in this paper, there are referred to as $inception_n (1 \leq n \leq 10)$.

3.3. Features from Gray-level co-occurrence matrices

“Hand-crafted” features were extracted from Gray-level co-occurence matrices (GLCM) [22] constructed using the designated ROIs. GLCMs in the four principal directions were constructed and then summed up to obtain a single direction invariant GLCM. A set of features [41], *i.e.* absolute value, entropy, contrast, energy, correlation, and homogeneity, computed using the GLCM for each ROI. These features were also concatenated into a high-dimensional feature vector (referred to as $GLCM_{6D}$) to be used for classification.

3.4. Classification

Subsequent to feature extraction, we performed a supervised learning step, using support vector machines (SVM), where the ROIs were classified as healthy or osteoarthritic. Briefly,

a SVM finds a linear separating hyperplane with the maximal margin in the feature space. The SVM implementation was taken from the Scikit-learn library[42]. As the performance obtained with the features extracted was optimal using the linear kernel, we have not explored additional higher dimensional kernels in this study. This also served to reduced the number of parameters to be optimized during cross-validation. The penalty term, C for the classifier was determined to be 100 based on empirical experiments.

To avoid overfitting we imposed the following restrictions during our data analysis steps. (1) We defined *non-overlapping* ROIs in the radial zone of the cartilage matrix in the PCI-CT images for each subject. This avoided overrepresentation of specific patterns from each subject. (2) We separated out 1 healthy and 1 diseased subject randomly, in a single iteration of the machine-learning step, for testing. The remaining 3 subjects were used for training the classifier (or fine-tuning the CaffeNet). In contrast to random splitting of the dataset into a training/test sets this ensures that ROIs from the same subject are never used towards training as well as testing, thereby preventing overfitting of the classifier and biasing the performance evaluation.

Performance of the classifier over the different iterations (training/test split) was evaluated using Receiver operating characteristics (ROC) curve. Here the SVM is used to produce probability scores for each class, which are calibrated using Platt scaling *i.e.* logistic regression on the SVMs scores, fit by an additional 5-fold cross-validation on the training data. These scores are then thresholded systematically to obtain a measure of true-positive rate (sensitivity) and false positive rate (1 − specificity), resulting in a ROC curve. The area under the ROC curve (AUC) is used as the evaluation metric for a particular feature set.

3.5. Visualization of CNN features

3.5.1. Activation of layers/neurons—Features obtained from CNNs can also be studied using appropriate visualization techniques. One such approach, is to visualize the activation of neurons during a forward pass through the network. In CNNs, individual neurons act as filters applied over the two spatial dimensions of the image, which in turn produces activations (or input) for subsequent layers. Such tools have been used previously to identify neuron which respond to specific patterns in the images such as text, flowers, faces etc [43]. Qualitative visualization of the features can help gain insight into CNNs especially with regards to distinguishing patterns of osteoarthritis in the cartilage matrix.

3.5.2. Dimension Reduction—We also explored the use of an unsupervised dimension reduction technique known as t-Distributed Stochastic Neighbor Embedding (t-SNE) [44], for analyzing the representative power of the features for distinguishing between healthy and osteoarthritic ROIs.

The central idea of Stochastic Neighbor Embedding (SNE) is to compute conditional probabilities representing similarities from Euclidean distances between high-dimensional vectors [44]. The mismatch between probabilities, once established in a high as well as low dimensional space, need to be minimized to obtain an accurate dimension-reduction. Although the technique produces good visualization of high-dimensional datasets, its application can be limited by the cost function, which is difficult to optimize, and the so-

called crowding problem[44]. t-SNE was introduced to overcome these limitations and it differs from SNE in two ways: (1) it uses a symmetrized version of the SNE cost function with simpler gradients and (2) it uses a Student-t distribution rather than a Gaussian to compute the similarity between two points in the low-dimensional space. More details pertaining to this algorithm and its cost function minimization can be found in [44], and a detailed review of the algorithm is available in[45, 46]. This technique has been shown to be particularly applicable for visualization of the high-dimensional data.

4. Results

4.1. Features from CaffeNet

For features extracted from a pre-trained CaffeNet, the best classification performance was obtained using the features extracted from the last convolutional layer ($conv_5$, mean AUC=0.91) and the last fully connected layer (fc_8, mean AUC=0.91). Interestingly, the commonly used first fully connected layer in such studies does not perform as well (AUC=0.81±0.17). The initial convolutional layers perform poorly at the classification task (Fig. 2).

Furthermore, fine-tuning (with appropriate training/test set separation) an improvement in classification performance was obtained (AUC 0.96 ± 0.07).

4.2. Features from Inception-v3

The performance of features from the Inception-v3 network are shown in Fig 3. In general features from all inception modules can accurately distinguish between the two classes (AUC > 0.95), with no significant differences in performance seen through the different layers.

4.3. Features from Gray-level co-occurrence matrices

Standard texture features extracted with the ROIs were also evaluated within the same cross validation scheme. Most GLCM derived texture features perform poorly (Table 2) with the exception of *Correlation* which produced a high AUC of (0.93±0.07). Interestingly, the 6-dimensional vector of GLCM features does not perform well at the classification.

4.4. Visualization of CNN features

4.4.1. Activation of layers/neurons—To perform qualitative visualization of the features produced by the datast we used the DeepVis toolbox [43] to study activations produced by the network. We noticed differential activation in layers that produced high classification performance. For example, in *conv5* layer we noticed specific neurons that mostly produced high activation for diseased samples (Fig. 4, red box). Similarly some neurons respond preferentially to healthy samples (Fig. 4, green box). Interestingly, layer *fc6* qualitatively produced similar activations for ROIs from both groups, however *fc7* did not (Supplementary Figures 1 & 2).

4.4.2. Dimension Reduction—We have also explored the visualization of features from both networks using t-SNE (Fig.5). The visualizations produced distinct clustering of

healthy and diseased ROIs, in-line with classification performance as obtained in the previous sections. We show here visualizations of features exhibiting best performance in CaffeNet (with and without fine tuning), namely $conv_5$ and fc_8. Given that features from Inception-v3 performs equally well, we included the last two inception modules for a comparative analysis.

5. Discussion

The applicability of Phase Contrast Imaging with Computed Tomography (PCI-CT) for visualizing structural details of the cartilage matrix with micrometer resolution has been demonstrated previously [1, 20]. This technique enables the visualization of differences in chondrocyte organization between healthy and osteoarthritic cartilage samples. In this study, we explored the application of deep transfer learning for characterizing the organizational patterns using multiple layers of convolutional neural networks, in a computer-aided diagnostics framework. Our results illustrate that features extracted from networks pretrained on non-medical imaging datasets can accurately classify chondrocyte patterns as normal or osteoarthritic with high accuracy.

Few studies have investigated the utility of machine learning in Phase contrast imaging data [47, 20, 21]. In previous works, we demonstrated the application of textural characterization of Phase contrast imaging data and its application in a computer aided diagnostic framework. We have shown that the use of texture features based on topological [20], statistical or geometric properties [48] in combination with machine learning can be used to distinguish healthy and osteoarthritic patterns in such images with high accuracy. These studies primarily focused on introducing novel texture features for characterizing the chondrocyte patterns observed in the PCI-CT images. Over the past few years focus has shifted from such “hand-crafted” features to those extracted from pretrained deep convolutional neural networks (CNN). In this study we intended to apply such deep learning techniques, not applied previously to PCI-CT data, and investigate the fidelity of features obtained for classification.

The current popularity of deep learning techniques stems from the ability of such networks to accurately capture patterns within images and hence produce high performance at various different tasks [24, 25, 39]. This is further emphasized through the application of transfer learning methods, wherein networks trained in a different setting can be adapted for a new task[29, 23]. These methods have recently gained popularity in medical imaging applications [49, 26, 50], where, networks trained on the ImageNet dataset have been used for the classification and detection of diseases [27, 28, 51, 52], often with improved level of performance over traditional, “hand-crafted”, features. We investigated the applicability of such methods on a dataset comprising of ROIs extracted from images of healthy and osteoarthritic cartilage acquired using PCI-CT.

It is widely accepted that in a deep learning network, the initial layers capture features that are generic (low-level features such as edges, wavelet filters etc.) and as we go deeper they get more specific for the task they are being trained on. Yosinski et al. [29] explored this aspect of transfer learning and noted that features from all layers are not easily transferable,

especially if, the datasets for two tasks are significantly different, which is the case in our study. We observe that the choice of layer from which specific features are extracted affects the overall results significantly. It is noteworthy that features extracted from the first fully connected layer (fc_6) of CaffeNet [34], which has been widely used in transfer learning studies does not perform as well as other layers. In fact, pooled features from the last convolutional layer ($conv_5$) and the last fully connected layer (fc_8) perform the best indicating adequate characterization of chondrocyte patterns through the high-level, "abstract" features captured in these (Fig. 2). The poor performance of features from the first convolutional layer ($conv_1$) before and after fine-tuning indicating that patterns of degeneration in osteoarthritic cartilage cannot be captured by the use of low-level features. This re-affirms that the choice of layer from CaffeNet for transfer learning should be made in a principled manner based on the target application, or in the case of using a smaller network such as CaffeNet should be reported for all layers as previously suggested [26]. When the CaffeNet was fine-tuned with our dataset, a significant improvement in performance was obtained (AUC (0.96 ± 0.07)). Hence, a significant improvement in the performance of convolutional neural networks can be obtained with a small amount of fine-tuning when the parameters (learning rate, momentum, initialization scheme etc.) are appropriately chosen. Post-fine tuning very high AUC values are obtained, indicating that some of the representations learnt, accurately capture changes in chondrocyte patterns during osteoarthritis.

Interestingly, the features from the inception modules of the Inception-v3 network [39, 53], all exhibit high classification performance (Fig. 3). Layers of this network, outperformed the CaffeNet suggesting that a more state-of-the art network provides an enhanced characterization and hence may not require fine-tuning of the network parameters for classification tasks. As performance obtained was fairly high (AUCs > 0.94) for all layers, we have not performed fine-tuning for this network. Although, variants of the AlexNet are still popular and continue to be widely used, it can be advocated that modern architectures such as Inception Networks or ResNets [54] could enhance performance of transfer learning methods.

It is worthwhile to mention here that although such high classification performances (AUC > 0.95) are atypical for traditional computer aided diagnostics studies, it is becoming more common place with the advent of deep learning [55, 47]. Applications of deep learning (as well as transfer-learning) based methods have in fact produced close to human performance across different domains of medical imaging [53, 56, 26]. This indicates that an exceptional characterization of images using CNNs which can substantially enhance computer aided-diagnostics. Researchers in computer vision have suggested the use of transfer learning to serve as potential benchmarks for evaluating newer methods [23]. To test this aspect and also compare the results obtained here with our prior work [20, 48] investigating textural characterization of phase contrast imaging data, we also computed "hand-crafted" features, derived from GLCM [22], on our dataset. Neither the individual features nor the 6 dimensional vector of features, however, perform as well as CNN features on the classification task (Table 2). This suggests the potential of CNN based methods for developing novel imaging based biomarkers for osteoarthritis with high-resolution imaging.

A common criticism of deep learning methods is the low interpretability of the representation(s) learned by the networks (beyond the first few layers) [29, 57]. Exploratory visualization techniques provide an alternative strategy to explore applicability of such features and ehance their representation [44, 58]. Hence apart from quantitative analysis of the characterization power of deep learning features through classification, we also have also applied two different visualization techniques for CNNs in a medical imaging setting. An interesting tool for visualization developed by Yosinski *et. al.* [43] has been used to show that when working with non-medical datasets, specific neurons within layers elicit preferential response to high-level features such as text, flower or faces. As the ROIs used in this study were annotated specifically in the radial zone of the matrix, such high-level features are not present. Using the DeepVis toolbox, we were however, able to identify neurons which tend to pick up clustered/checkered patterns (in the ImageNet dataset) seem to respond more to ROIs extracted from normal subjects compared to those from OA (Fig 4). The visualization of activations in such a manner gives an intuitive understanding of the patterns picked up by CNNs can improve the interpretability of features. It can provide valuable insights for transfer learning studies to better apply and adapt networks trained on non-medical datasets and also be used to motivate exact the choice of features in a medical imaging task. For example, neurons responding to checkered patterns are more applicable for classifying PCI-CT data than say a neuron that responds to text. We anticipate the use of such a visualization technique could potentially contribute further in more complicated tasks such as organ segmentation, lesion detection etc. as larger medical imaging datasets become available.

We also visualized the features using a dimension reduction technique, t-SNE [44], which has been shown to be effective in visualizations of high-dimensional data. For features from off-the-shelf CaffeNet we notice the a slight overlap between the healthy and diseased subject groups, which is interestingly reduced post fine-tuning (Fig. 5). Similar to our quantitative analysis, features from the subject groups, Inception-v3 network, cluster together although some overlap is seen in features from some layers. We would however like to point out that classification performance using SVMs and cluster separation in using t-SNE need not always produce perfect agreement. It should also be noted that t-SNE is an *unsupervised* dimension reduction approach that can provide valuable insight regarding the separability of the the data. We see that reducing the dimension of features using t-SNE produces a good separation of the two groups, with most features (Supplementary Figure 3). This is indicative of the *possibility* of achieving good classification results, though a one to one correspondence between the two techniques cannot be expected in all scenarios, especially as the SVM is a *supervised* learning algorithm that uses training examples as support vectors for defining the decision boundary. Such a discrepancy can be observed, in the t-SNE visualizations obtained for features from $inception_4$, $inception_6$, & $inception_7$, wherein the clusters are not separable however using the features with SVM still produces a high classification accuracy.

Although our results are promising, we would like to acknowledge a few limitations of this study, in the current form. Firstly, the specimens used for imaging were obtained from a small number (five) of patients. Thereby, to avoid overfitting, we have introduced an additional restriction for ensuring strict separation of training and testing data, by randomly

assigning all ROIs extracted from one healthy and osteoarthritic case to testing set in each iteration. It, however, remains a possibility that the ROIs extracted from the two classes (healthy and osteoarthritic) could be over-represented due to the limited variations of patterns found in these subjects. Additionally, although OA related degeneration is progressive, we have not defined any healthy ROIs in the diseased samples to primarily avoid ambiguous ROI definition and prevent class imbalance. In future studies as the availability of PCI-CT systems increases, we aim to include more patients to perform a more robust evaluation of the methods proposed here, considering a full segmentation of a sample into healthy and diseased regions. The methodology followed in this study however allows the analysis of chondrocyte patterns at an unprecedented resolution for imaging studies. It has also enabled us to perform a detailed investigation into the feature sets obtained from different layers of CNNs and study their applicability in a computer-aided diagnostics task. Secondly, we also note a practical limitation with the imaging setup used in this study concerning the reliance of the imaging technique on synchrotron radiation. The use of a stationary radiation source restricts PCI-CT imaging to ex-*vivo* specimens, thereby further restricting the amount of imaging data which could be acquired. In this regard, ongoing research investigating alternatives strategies for improving of PCI methods show significant promise for enabling in-*vivo* and clinical imaging.

6. Conclusion

This study shows the applicability of deep transfer learning techniques to classify healthy and osteoarthritic chondrcyte patterns acquired from PCI-CT imaging of the human patellar cartilage. We explored the utility of feature representations extracted from two different convolutional networks: a simpler and widely used network, CaffeNet, as well as a network with more advanced architecture, Inception-v3. Our results show that, features extracted from the last convolutional layer and last fully connected layer of CaffeNet perform significantly better than other layers, suggesting that an informed choice regarding selection of layer for feature extraction is critical for achieving good performance. Although, we have used a smaller dataset for this study, we have shown that fine-tuning, when applied appropriately, can aid in improving the performance of such networks. Features extracted from the modules of the Inception-v3 network produce excellent classification performance even without fine-tuning. Thus, there is a potential for using such deep-transfer learning approaches, for detecting the presence osteoarthritis, in a computer-aided diagnosis framework. However, larger studies need to be conducted in order to further validate the clinical plausibility of such methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was funded in part by the National Institute of Health (NIH) Award R01-DA-034977, the HarryW. Fischer Award of the University of Rochester, 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, the Center for Emerging and Innovative Sciences (CEIS), a NYSTAR-designated Center for Advanced Technology. This work was performed as a practice quality improvement (PQI) project for

maintenance of certificate (MOC) of Axel Wismüller's American Board of Radiology (ABR) certification. The authors would like to thank the ESRF for providing the experimental facilities and the ESRF ID17 team for assistance in operating the facilities.

References

1. Coan P, Bamberg F, Diemoz PC, Bravin A, Timpert K, Mützel E, Raya JG, Adam-Neumair S, Reiser MF, Glaser C. Characterization of osteoarthritic and normal human patella cartilage by computed tomography x-ray phase-contrast imaging: a feasibility study. Investigative radiology. 2010; 45(7):437–444. [PubMed: 20479648]
2. Snigirev A, Snigireva I, Kohn V, Kuznetsov S, Schelokov I. On the possibilities of x-ray phase contrast microimaging by coherent high-energy synchrotron radiation. Review of scientific instruments. 1995; 66(12):5486–5492.
3. Davis T, Gao D, Gureyev T, Stevenson A, Wilkins S, et al. Phase-contrast imaging of weakly absorbing materials using hard x-rays. Nature. 1995; 373(6515):595–598.
4. Takeda T, Momose A, Itai Y, Jin W, Hirano K. Phase-contrast imaging with synchrotron x-rays for detecting cancer lesions. Academic radiology. 1995; 2(9):799–803. [PubMed: 9419642]
5. Keyrilainen J, Fernández M, Karjalainen-Lindsberg M-L, Virkkunen P, Leidenius M, Von Smitten K, Sipila P, Fiedler S, Suhonen H, Suortti P, et al. Toward high-contrast breast ct at low radiation dose 1. Radiology. 2008; 249(1):321–327. [PubMed: 18796684]
6. Schneider T, Coan P, Habs D, Reiser M. application of brilliant xrays in mammography. development and perspectives of phase contrast techniques. Der Radiologe. 2008; 48(4):345–350. [PubMed: 18311554]
7. Coan P, Mollenhauer J, Wagner A, Muehleman C, Bravin A. Analyzer-based imaging technique in tomography of cartilage and metal implants: a study at the esrf. European journal of radiology. 2008; 68(3):S41–S48. [PubMed: 18584983]
8. Muehleman C, Majumdar S, Issever AS, Arfelli F, Menk R-H, Rigon L, Heitner G, Reime B, Metge J, Wagner A, et al. X-ray detection of structural orientation in human articular cartilage. Osteoarthritis and cartilage. 2004; 12(2):97–105. [PubMed: 14723869]
9. Chapman D, Thomlinson W, Johnston R, Washburn D, Pisano E, Gmür N, Zhong Z, Menk R, Arfelli F, Sayers D. Diffraction enhanced x-ray imaging. Physics in medicine and biology. 1997; 42(11):2015. [PubMed: 9394394]
10. Bravin A. Exploiting the x-ray refraction contrast with an analyser: the state of the art. Journal of Physics D: Applied Physics. 2003; 36(10A):A24.
11. Connor DM, Benveniste H, Dilmanian FA, Kritzer MF, Miller LM, Zhong Z. Computed tomography of amyloid plaques in a mouse model of alzheimer's disease using diffraction enhanced imaging. Neuroimage. 2009; 46(4):908–914. [PubMed: 19303447]
12. Braun HJ, Gold GE. Diagnosis of osteoarthritis: imaging. Bone. 2012; 51(2):278–288. [PubMed: 22155587]
13. Bashir A, Gray M, Boutin R, Burstein D. Glycosaminoglycan in articular cartilage: in vivo assessment with delayed gd(DTPA)(2-)-enhanced MR imaging. Radiology. 1997; 205(2):551–558. [PubMed: 9356644]
14. Reddy R, Li S, Noyszewski E, Kneeland J, Leigh J. In vivo sodium multiple quantum spectroscopy of human articular cartilage. Magnetic Resonance in Medicine. 1997; 38(2):207–214. [PubMed: 9256099]
15. Stahl R, Luke A, Li X, Carballido-Gamio J, Ma C, Majumdar S, Link T. 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. 2009; 19(1):132–143. [PubMed: 18709373]
16. Schmitt B, Zbyn S, Stelzeneder D, Jellus V, Paul D, Lauer L, Bachert P, Trattnig S. Cartilage quality assessment by using glycosaminoglycan chemical exchange saturation transfer and 23 Na MR imaging at 7 T. Radiology. 2011; 260(1):257–264. [PubMed: 21460030]
17. Sophia Fox AJ, Bedi A, Rodeo SA. The basic science of articular cartilage: structure, composition, and function. Sports health. 2009; 1(6):461–468. [PubMed: 23015907]

18. Sandell LJ, Aigner T. Articular cartilage and changes in arthritis: cell biology of osteoarthritis. Arthritis Research & Therapy. 2001; 3(2):107.
19. Poole CA. Articular cartilage chondrons: form, function and failure. The Journal of Anatomy. 1997; 191(1):1–13. [PubMed: 9279653]
20. Nagarajan MB, Coan P, Huber MB, Diemoz PC, Glaser C, Wismüller A. Computer-aided diagnosis in phase contrast imaging x-ray computed tomography for quantitative characterization of ex vivo human patellar cartilage. IEEE Transactions on Biomedical Engineering. 2013; 60(10): 2896–2903. [PubMed: 23744660]
21. Abidin AZ, Nagarajan MB, Checefsky WA, Coan P, Diemoz PC, Hobbs SK, Huber MB, Wismüller A. Volumetric characterization of human patellar cartilage matrix on phase contrast x-ray computed tomography. SPIE Medical Imaging, International Society for Optics and Photonics. 2015:94171F–94171F.
22. Haralick RM. Statistical and structural approaches to texture. Proceedings of the IEEE. 1979; 67(5):786–804.
23. Sharif Razavian, A., Azizpour, H., Sullivan, J., Carlsson, S. Cnn features off-the-shelf: an astounding baseline for recognition; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops; 2014. p. 806-813.
24. Donahue J, Jia Y, Vinyals O, Hoffman J, Zhang N, Tzeng E, Darrell T. Decaf: A deep convolutional activation feature for generic visual recognition. ICML. 2014:647–655.
25. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015; 521(7553):436–444. [PubMed: 26017442]
26. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI. A survey on deep learning in medical image analysis. arXiv preprint arXiv:1702.05747.
27. Bar, Y., Diamant, I., Wolf, L., Lieberman, S., Konen, E., Greenspan, H. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE; 2015. Chest pathology detection using deep learning with non-medical training; p. 294-297.
28. van Ginneken, B., Setio, AA., Jacobs, C., Ciompi, F. 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI). IEEE; 2015. Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans; p. 286-289.
29. Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? Advances in neural information processing systems. 2014:3320–3328.
30. Schlegl, T., Ofner, J., Langs, G. International MICCAI Workshop on Medical Computer Vision. Springer; 2014. Unsupervised pre-training across image domains improves lung tissue classification; p. 82-93.
31. Pritzker K, Gay S, Jimenez S, Ostergaard K, Pelletier J-P, Revell P, Salter D, Van den Berg W. Osteoarthritis cartilage histopathology: grading and staging. Osteoarthritis and cartilage. 2006; 14(1):13–29. [PubMed: 16242352]
32. Ingal V, Beliaevskaya E. X-ray plane-wave topography observation of the phase contrast from a non-crystalline object. Journal of Physics D: Applied Physics. 1995; 28(11):2314.
33. Coan P, Peterzol A, Fiedler S, Ponchut C, Labiche JC, Bravin A. Evaluation of imaging performance of a taper optics ccdfrelon'camera designed for medical imaging. Journal of synchrotron radiation. 2006; 13(3):260–270. [PubMed: 16645252]
34. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T. Proceedings of the 22nd ACM international conference on Multimedia. ACM; 2014. Caffe: Convolutional architecture for fast feature embedding; p. 675-678.
35. Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. arXiv preprint arXiv:1512.00567.
36. Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems. 2012:1097–1105.
37. Scherer D, Müller A, Behnke S. Evaluation of pooling operations in convolutional architectures for object recognition. Artificial Neural Networks–ICANN. 2010; 2010:92–101.

38. Boureau, Y-L., Ponce, J., LeCun, Y. A theoretical analysis of feature pooling in visual recognition; Proceedings of the 27th international conference on machine learning (ICML-10); 2010. p. 111-118.
39. Lin M, Chen Q, Yan S. Network in network. arXiv preprint arXiv:1312.4400.
40. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A. Going deeper with convolutions; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2015. p. 1-9.
41. Anys H, He D. Evaluation of textural and multipolarization radar features for crop classification. IEEE Transactions on Geoscience and Remote Sensing. 1995; 33(5):1170–1181.
42. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research. 2011; 12:2825–2830.
43. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding neural networks through deep visualization. arXiv preprint arXiv:1506.06579.
44. Maaten, Lvd, Hinton, G. Visualizing data using t-sne. Journal of Machine Learning Research. 2008; 9(Nov):2579–2605.
45. Jamieson AR, Giger ML, Drukker K, Li H, Yuan Y, Bhooshan N. Exploring nonlinear feature space dimension reduction and data representation in breast cadx with laplacian eigenmaps and t-sne. Medical physics. 2010; 37(1):339–351. [PubMed: 20175497]
46. Bunte K, Hammer B, Villmann T, Biehl M, Wismüller A. Neighbor embedding xom for dimension reduction and visualization. Neurocomputing. 2011; 74(9):1340–1350.
47. Mao, Y., Yin, Z. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2016. A hierarchical convolutional neural network for mitosis detection in phase-contrast microscopy images; p. 685-692.
48. Nagarajan MB, Coan P, Huber MB, Diemoz PC, Glaser C, Wismüller A. Computer-aided diagnosis for phase-contrast x-ray computed tomography: Quantitative characterization of human patellar cartilage with high-dimensional geometric features. Journal of digital imaging. 2014; 27(1):98–107. [PubMed: 24043594]
49. Xu, Y., Mo, T., Feng, Q., Zhong, P., Lai, M., Eric, I., Chang, C. Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on. IEEE; 2014. Deep learning of feature representation with multiple instance learning for medical image analysis; p. 1626-1630.
50. Lopes U, Valiati J. Pre-trained convolutional neural networks as feature extractors for tuberculosis detection. Computers in Biology and Medicine. 2017; 89:135–143. [PubMed: 28800442]
51. Shin H-C, Roth HR, Gao M, Lu L, Xu Z, Nogues I, Yao J, Mollura D, Summers RM. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. IEEE transactions on medical imaging. 2016; 35(5):1285–1298. [PubMed: 26886976]
52. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Computers in Biology and Medicine.
53. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S. Dermatologist-level classification of skin cancer with deep neural networks. Nature. 2017; 542(7639):115–118. [PubMed: 28117445]
54. He, K., Zhang, X., Ren, S., Sun, J. Deep residual learning for image recognition; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. p. 770-778.
55. Bejnordi BE, Litjens G, Timofeeva N, Otte-Höller I, Homeyer A, Karssemeijer N, van der Laak JA. Stain specific standardization of whole-slide histopathological images. IEEE transactions on medical imaging. 2016; 35(2):404–415. [PubMed: 26353368]
56. Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. Jama. 2016; 316(22):2402–2410. [PubMed: 27898976]
57. Long, M., Cao, Y., Wang, J., Jordan, M. Learning transferable features with deep adaptation networks; International Conference on Machine Learning; 2015. p. 97-105.

58. Cho K, Van Merriënboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, Bengio Y. Learning phrase representations using rnn encoderdecoder for statistical machine translation. arXiv preprint arXiv:1406.1078.

- Phase Contrast Imaging allows visualization of osteoarthritic changes in the patellar cartilage
- Features from pre-trained CNNs can be used characterize healthy and diseased patterns
- Features from Inception-v3 perform better than CaffeNet and GLCM at the classification task
- Fine-tuning with a small dataset can improve classifier performance
- Visualization techniques can help further substantiate the characterization obtained

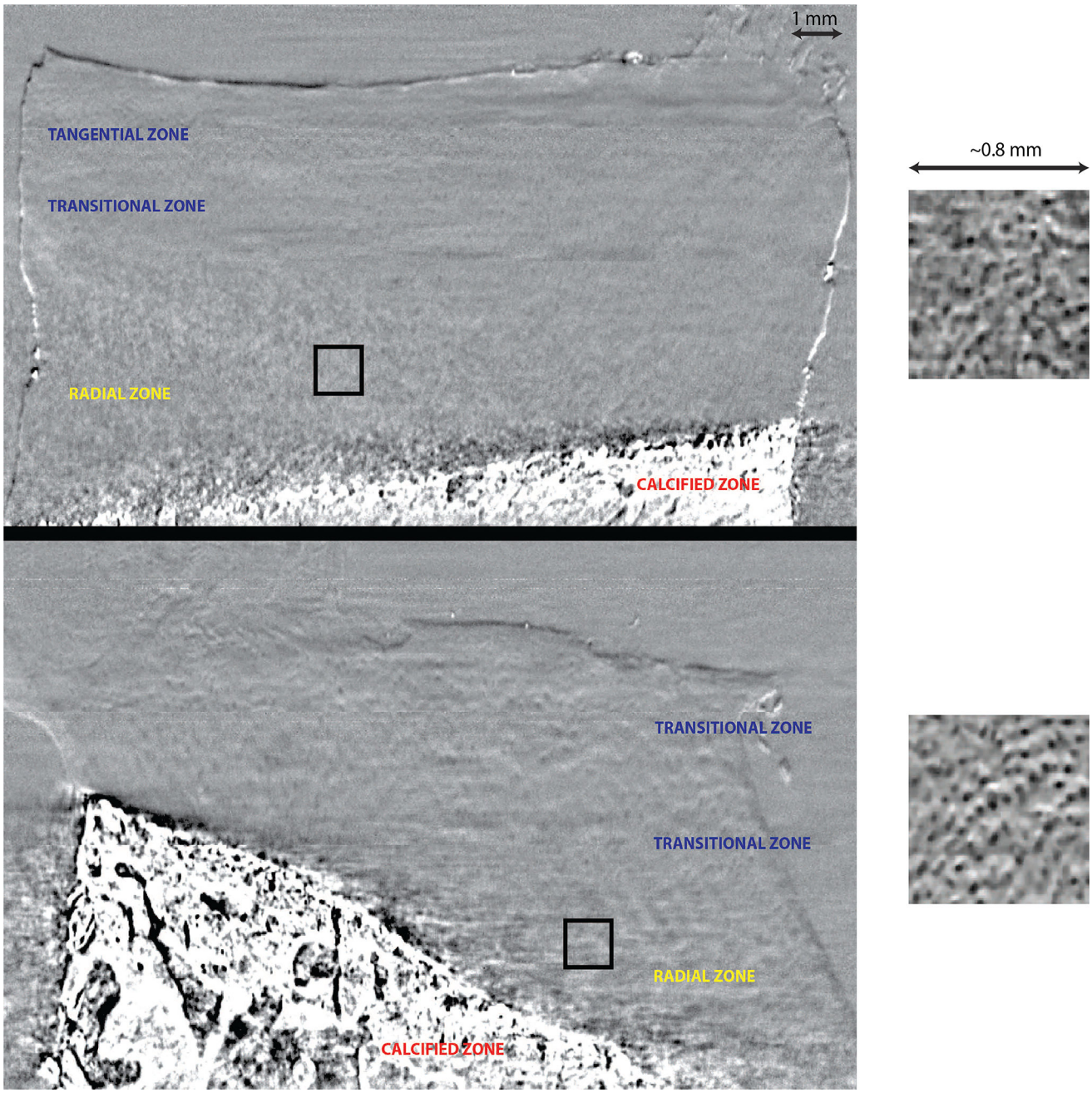

Figure 1. Cross-sectional view of single slice of articular cartilage as obtained using PCI-CT

The presence of an orderly zonal architecture (tangential, transitional and radial) can be clearly visualized via PCI-CT imaging, particularly in healthy samples **(Top)**. This zone-specific organization of chondrocytes gradually degrades during osteoarthritis **(Bottom)** and is instead replaced by a more generalized clustering of cells throughout the matrix accompanied by a loss of clear zonal separation. The black boxes indicate an example ROI definition in the radial zone extracted from both the groups. Sample ROIs from both the classes, used for the analysis, are magnified and shown on the right.

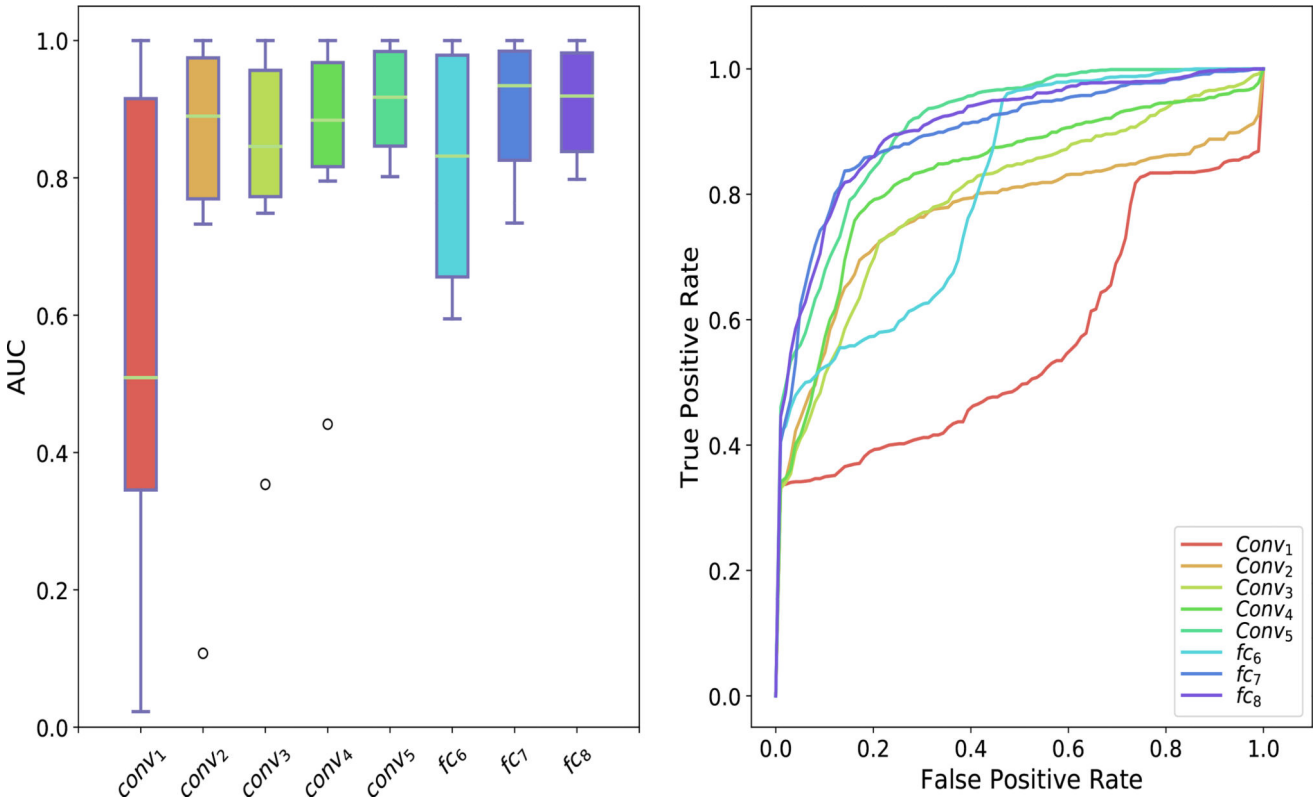

Figure 2. Comparison of classification performance for features extracted from different layers of a pre-trained CaffeNet

(Left) Boxplots representing the performance of features from different layers over multiple iterations. Each of the colored regions indicates the 25^{th} and the 75^{th} percentiles and the central line corresponds to the median AUC value across different test/train splits. Whiskers extend up to 1.5 times the interquartile range and the small circles, when seen, represents statistical outliers. Some AUC values for the first few layers were < 0.5, which representing reversal in class estimation. **(Right)** The corresponding mean ROC curves generated with features from different layers. Representations extracted from the last convolutional as well as the last fully connected layer perform best at the classification task.

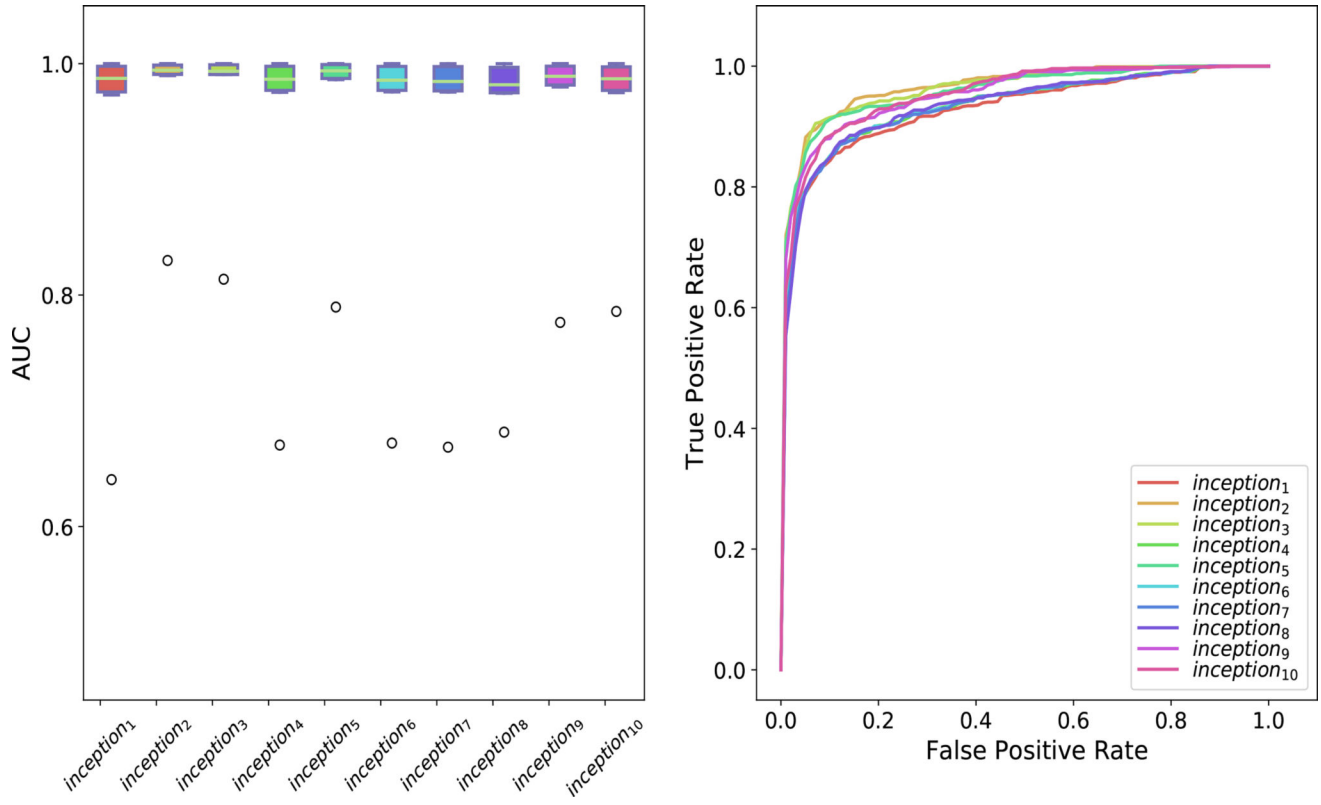

Figure 3. Comparison of classification performance for features extracted from different inception modules of a pre-trained Inception-v3 network

(Left) Boxplots representing the performance of features from different layers over multiple iterations. Each of the colored regions indicates the 25^{th} and the 75^{th} percentiles and the central line corresponds to the median AUC value across different test/train splits. Whiskers extend up to 1.5 times the interquartile range and the small circles, when seen, represents statistical outliers. **(Right)** The corresponding mean ROC curves generated with features from different modules. Features from all modules perform well at the classification task.

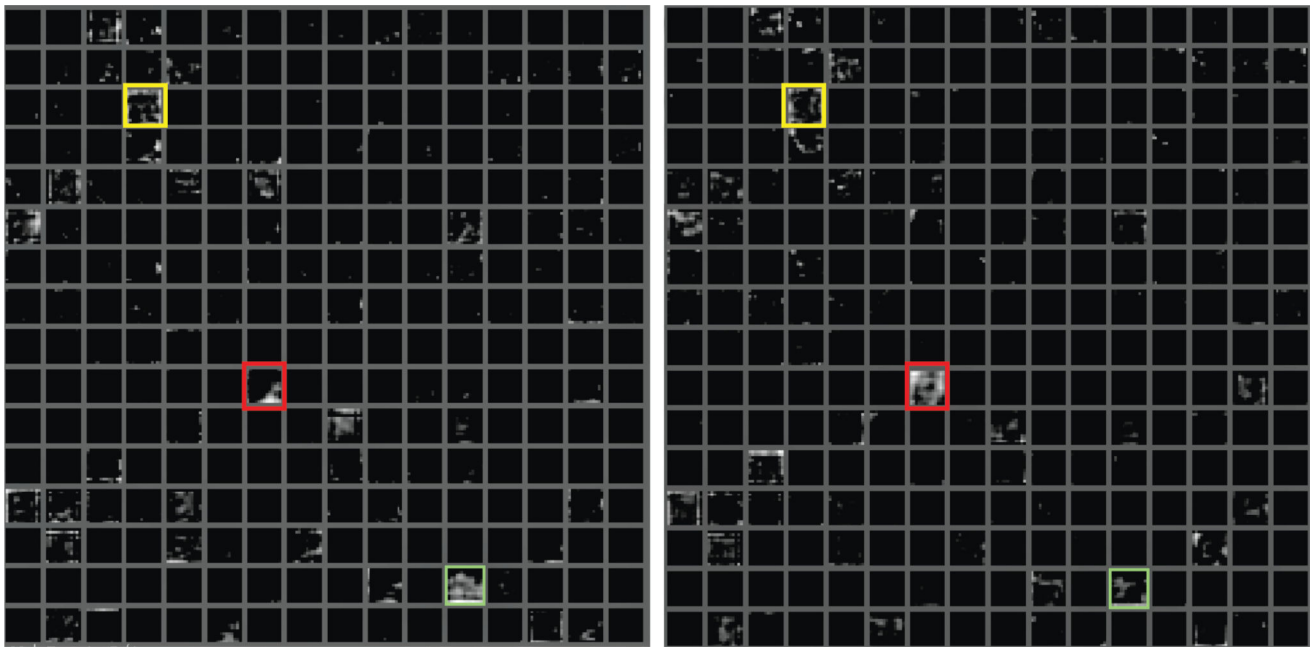

Figure 4. Differential activation of neurons to healthy (left) and osteoarthritic (right) samples in all 256 neurons in the *conv5* layer of CaffeNet visualized using the DeepVis toolbox

It was noticed that neuron 150 (red box) generally produced high activation in response to OA samples. In contrast neuron 221 (green box) produced an opposite response. Many neurons produced similar responses in this layer, for ex. neuron 36 (yellow box). Exploratory visualization techniques can motivate the choice of specific layers for extraction of CNN features. These visualizations have been adapted from the DeepVis toolbox.[43]

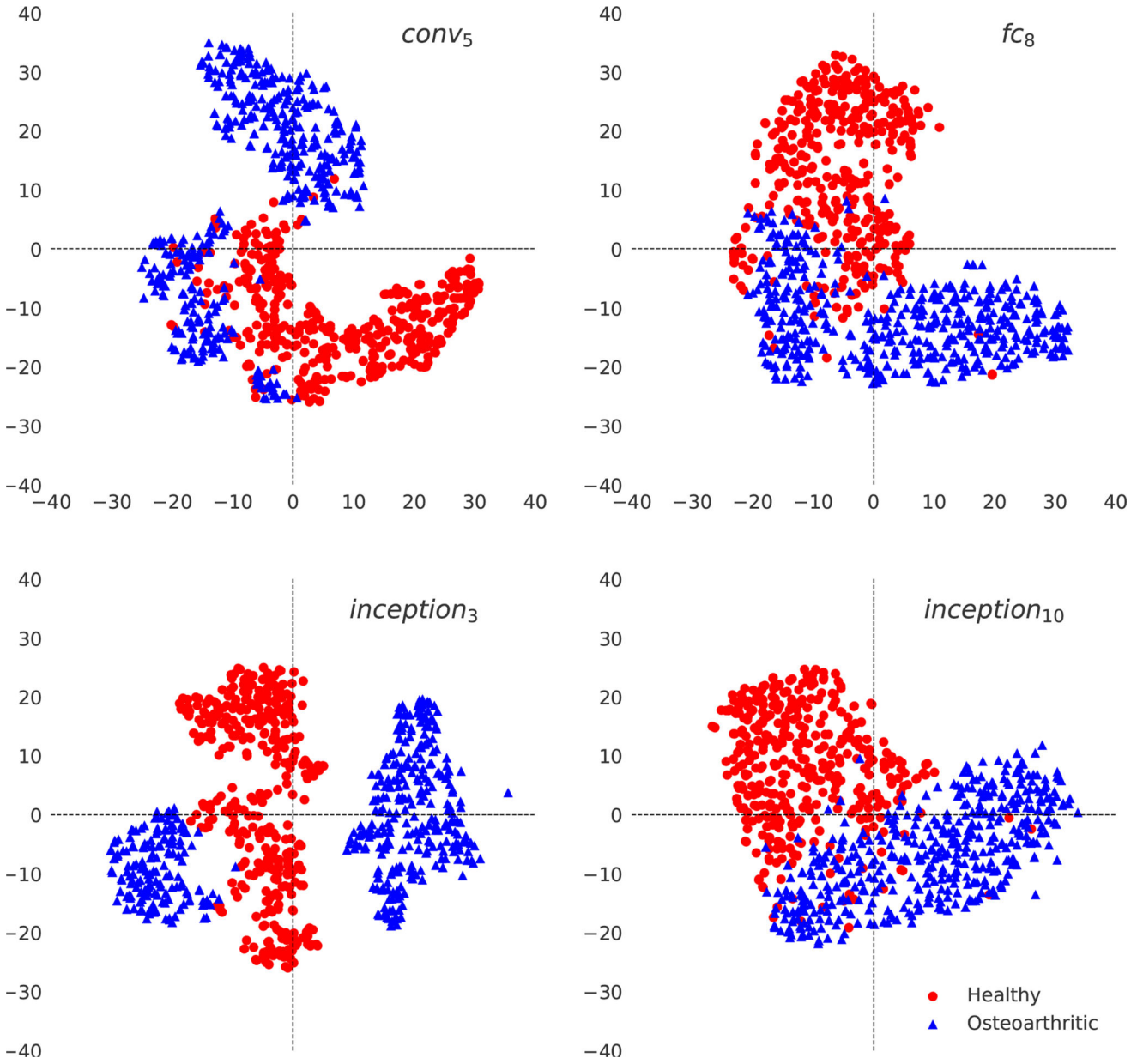

Figure 5. Visualization of high-dimensional features obtained using t-SNE dimension reduction

The visualizations of the features, that performed best at the classification task, are compared. We notice a clear distinction between healthy (red) and diseased (blue) clusters re-iterating that such features capture adequate information for distinguishing between the two classes.

Table 1

The architecture of Inception-v3 Network. The three dimensions of "input size" are channel dimension, width and height respectively. Features were extracted from the final 10 inception modules.

Type of Layer	patch size / stride	input size
conv	3 × 3/2	3 × 299 × 299
conv_1	3 × 3/1	32 × 149 × 149
conv_2	3 × 3/1	32 × 147 × 147
pool	3 × 3/2	64 × 147 × 147
conv_3	1 × 1/1	64 × 73 × 73
conv_4	3 × 3/1	80 × 73 × 73
pool_1	3 × 3/2	192 × 71 × 71
3× inceptionA	Figure 5 of [35]	192 × 35 × 35
inceptionB	Figure 10 of [35]	288 × 35 × 35
4× inceptionC	Figure 6 of [35]	768 × 17 × 17
inceptionD	Figure 10* of [35]	768 × 17 × 17
2×inceptionE	Figure 7 of [35]	1280 × 8 × 8
Global pool	—	2048 × 8 × 8
fc (Softmax)	—	2048 × 1 × 1

Table 2

Comparison of AUC (mean ± standard deviation) values obtained for the different feature sets used in this study. The best performing features within each class are **highlighted**.

CaffeNe	AUC	Inception-v3	AUC	GLCM	AUC
$conv_1$	0.56 ± 0.36	$inception_1$	0.93 ± 0.13	*Absolute Value*	0.78 ± 0.12
$conv_2$	0.77 ± 0.31	$inception_2$	0.96 ± 0.06	*Entropy*	0.57 ± 0.04
$conv_3$	0.80 ± 0.22	$inception_3$	0.96 ± 0.07	*Contrast*	0.78 ± 0.13
$conv_4$	0.83 ± 0.19	$inception_4$	0.94 ± 0.12	*Energy*	0.59 ± 0.06
$conv_5$	0.91 ± 0.08	$inception_5$	0.96 ± 0.07	***Correlation***	0.93 ± 0.07
$fc6_6$	0.81 ± 0.17	$inception_6$	0.94 ± 0.12	*Homogeneity*	0.76 ± 0.12
fc_7	0.90 ± 0.10	$inception_7$	0.94 ± 0.12	$GLCM_{6D}$	0.83 ± 0.15
fc_8	0.91 ± 0.08	$inception_8$	0.94 ± 0.11		
		$inception_9$	0.96 ± 0.08		
FT	0.96 ± 0.07	$inception_{10}$	0.95 ± 0.07		

Comput Biol Med. Author manuscript; available in PMC 2019 April 01.